Classification

ML - Supervised Learning

Dan Calderone - Win22

Classification

OUTPUTS

(Dependent Variables)

$$\gamma_t = \mathrm{sgn}(\theta^T x_t - \theta_0)$$

INPUTS

(Independent Variables)

$$\begin{bmatrix} \boxed{\begin{matrix}\gamma_{00} \\ \gamma_{10} \\ \gamma_{20} \\ \gamma_{30} \\ \gamma_{40} \\ \vdots \\ \gamma_{T0}\end{matrix}} & \begin{matrix}\cdots \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ \\ \cdots\end{matrix} & \begin{matrix}\gamma_{0m'} \\ \gamma_{1m'} \\ \gamma_{2m'} \\ \gamma_{3m'} \\ \gamma_{4m'} \\ \vdots \\ \gamma_{Tm'}\end{matrix} & \begin{matrix}y_{00} \\ y_{10} \\ y_{20} \\ y_{30} \\ y_{40} \\ \vdots \\ y_{T0}\end{matrix} & \begin{matrix}\cdots \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ \\ \cdots\end{matrix} & \begin{matrix}y_{0m} \\ y_{1m} \\ y_{2m} \\ y_{3m} \\ y_{4m} \\ \vdots \\ y_{Tm}\end{matrix} \end{bmatrix} \Leftarrow \boxed{f\left[\left(\begin{matrix} x_{00} & \cdots & x_{0n} \\ x_{10} & \cdots & x_{1n} \end{matrix}\right)\right.} \begin{matrix} x_{20} & \cdots & x_{2n} \\ x_{30} & \cdots & x_{3n} \\ x_{40} & \cdots & x_{4n} \\ \vdots & & \vdots \\ x_{T0} & \cdots & x_{Tn} \end{matrix} \Bigg] \quad \begin{bmatrix} \xi_{00} & \cdots & \xi_{0n'} \\ \xi_{10} & \cdots & \xi_{1n'} \\ \xi_{20} & \cdots & \xi_{2n'} \\ \xi_{30} & \cdots & \xi_{3n'} \\ \xi_{40} & \cdots & \xi_{4n'} \\ \vdots & & \vdots \\ \xi_{T0} & \cdots & \xi_{Tn'} \end{bmatrix}$$

"Binary classifier"

Support Vector Machine (SVM)

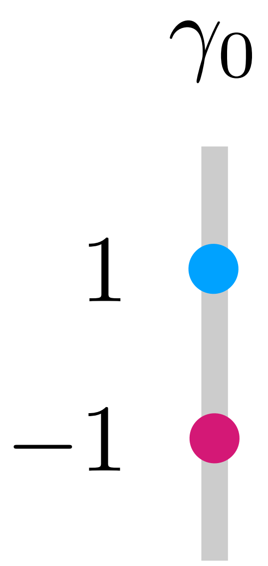
COST:

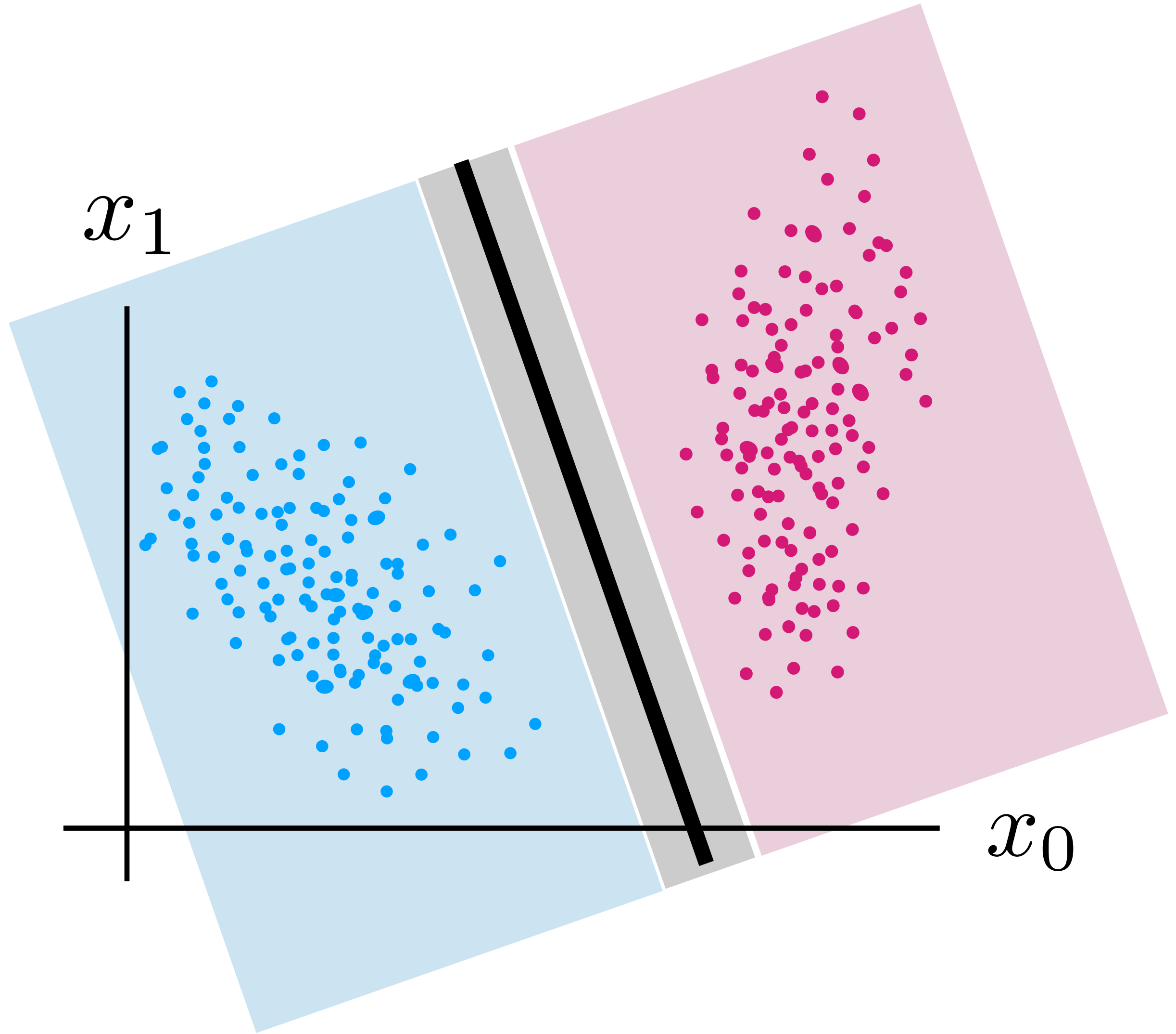
$$\min_{\theta} \quad ||\theta_{1:n}||_2^2$$

$$\text{s.t.} \quad \gamma_t(\theta_{1:n}^T x_t - \theta_0) \geq 1$$

Hard boundary

"Binary classifier"





Classification

OUTPUTS

(Dependent Variables)

$$\gamma_t = \text{sgn}(\theta^T x_t - \theta_0)$$

INPUTS

(Independent Variables)

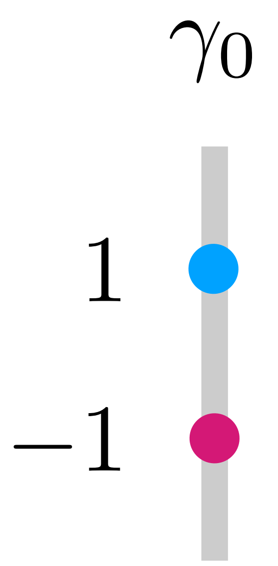
$$\begin{bmatrix} \boxed{\begin{matrix} \gamma_{00} \\ \gamma_{10} \\ \gamma_{20} \\ \gamma_{30} \\ \gamma_{40} \\ \vdots \\ \gamma_{T0} \end{matrix}} & \begin{matrix} \cdots \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ \\ \cdots \end{matrix} & \begin{matrix} \gamma_{0m'} \\ \gamma_{1m'} \\ \gamma_{2m'} \\ \gamma_{3m'} \\ \gamma_{4m'} \\ \vdots \\ \gamma_{Tm'} \end{matrix} & \begin{matrix} y_{00} \\ y_{10} \\ y_{20} \\ y_{30} \\ y_{40} \\ \vdots \\ y_{T0} \end{matrix} & \begin{matrix} \cdots \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ \\ \cdots \end{matrix} & \begin{matrix} y_{0m} \\ y_{1m} \\ y_{2m} \\ y_{3m} \\ y_{4m} \\ \vdots \\ y_{Tm} \end{matrix} \end{bmatrix} \Leftarrow \boxed{f\left[\left(\begin{matrix} x_{00} & \cdots & x_{0n} \\ x_{10} & \cdots & x_{1n} \end{matrix} \right)\right.} \begin{matrix} \\ \\ x_{20} & \cdots & x_{2n} \\ x_{30} & \cdots & x_{3n} \\ x_{40} & \cdots & x_{4n} \\ \vdots & & \vdots \\ x_{T0} & \cdots & x_{Tn} \end{matrix} \Bigg] \begin{bmatrix} \xi_{00} & \cdots & \xi_{0n'} \\ \xi_{10} & \cdots & \xi_{1n'} \\ \xi_{20} & \cdots & \xi_{2n'} \\ \xi_{30} & \cdots & \xi_{3n'} \\ \xi_{40} & \cdots & \xi_{4n'} \\ \vdots & & \vdots \\ \xi_{T0} & \cdots & \xi_{Tn'} \end{bmatrix}$$

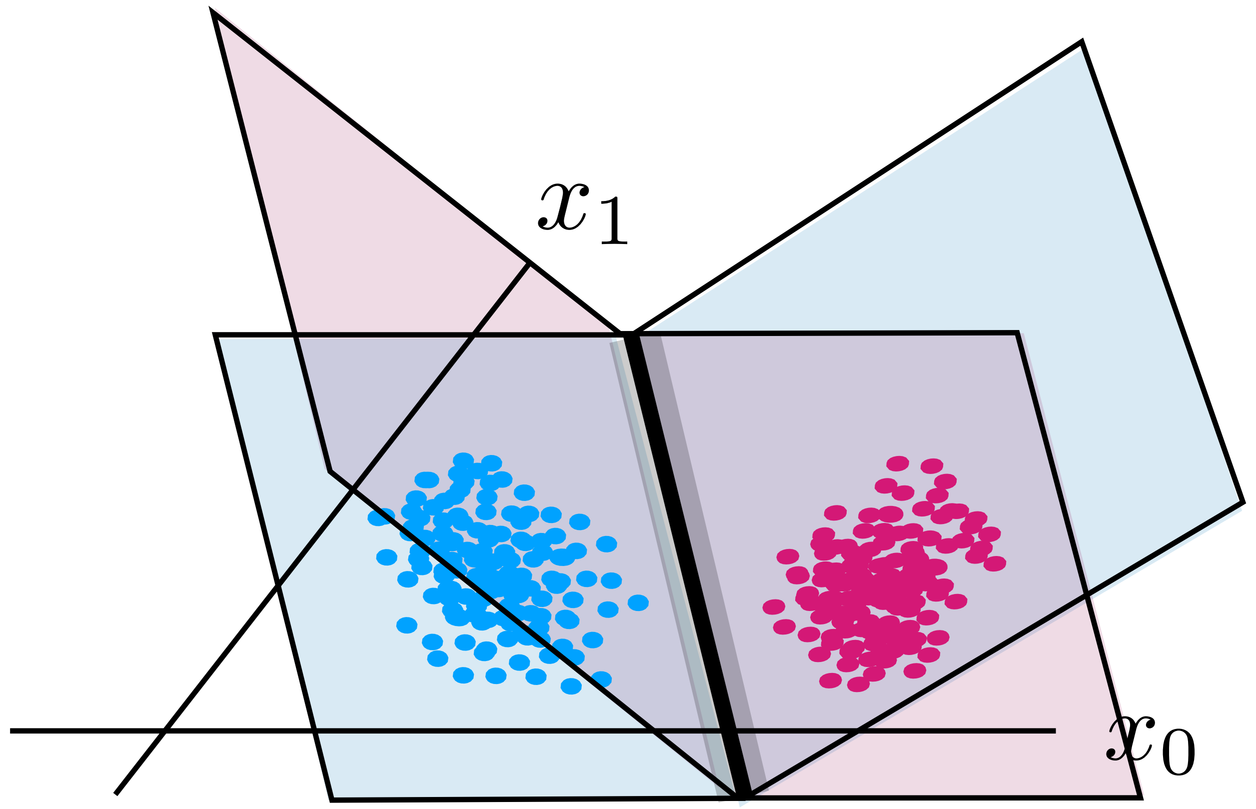
"Binary classifier"

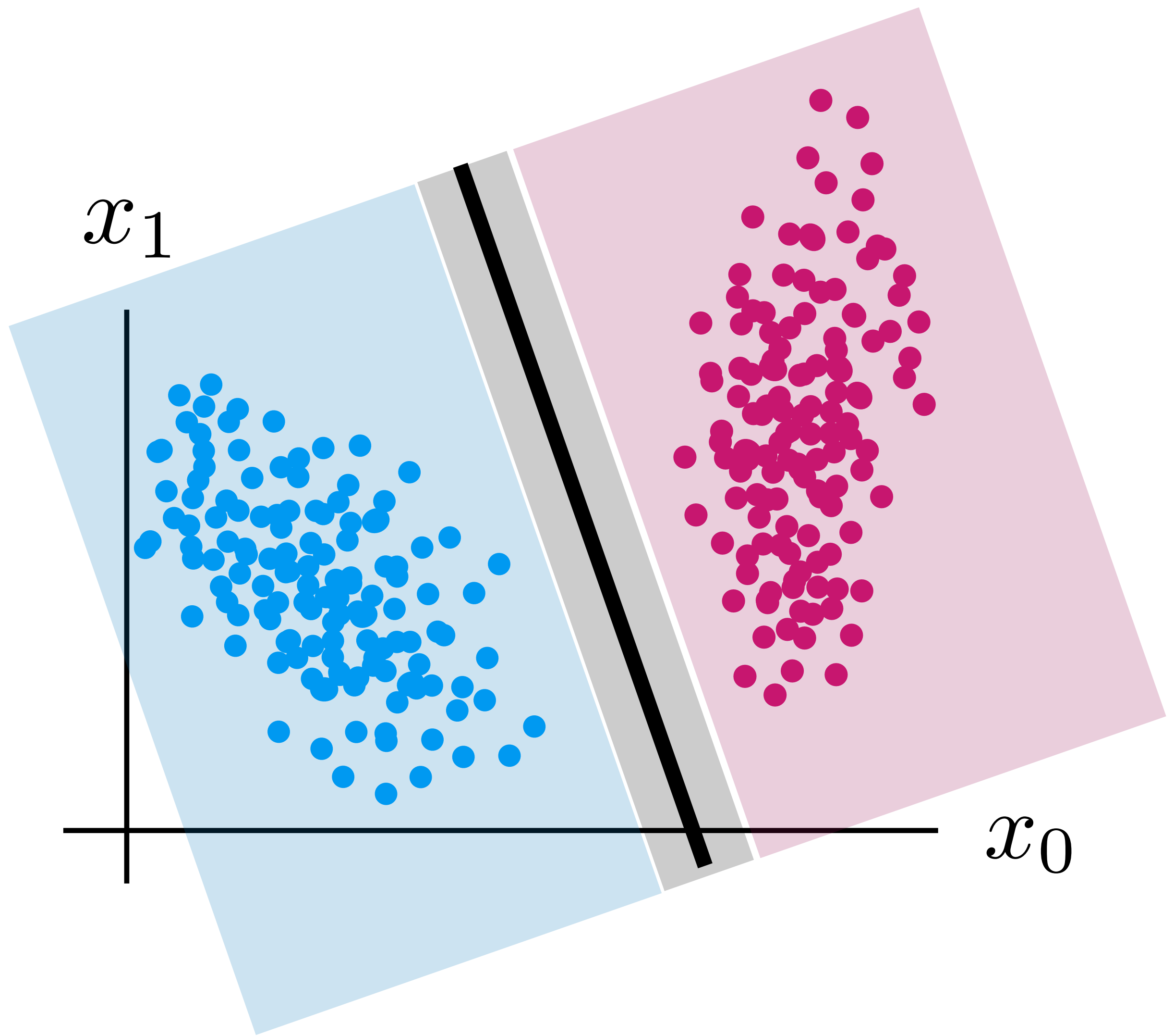
Support Vector Machine (SVM)

COST: $$\min_{\theta} \; \lambda ||\theta_{1:n}||_2^2 + \frac{1}{T}\sum_t \max\{0, 1 - \gamma_t(\theta_{1:n}^T x_t - \theta_0)\}$$

Soft boundary







Classification

OUTPUTS

(Dependent Variables)

$$\gamma_t = \mathrm{sgn}(\theta^T x_t - \theta_0)$$

INPUTS

(Independent Variables)

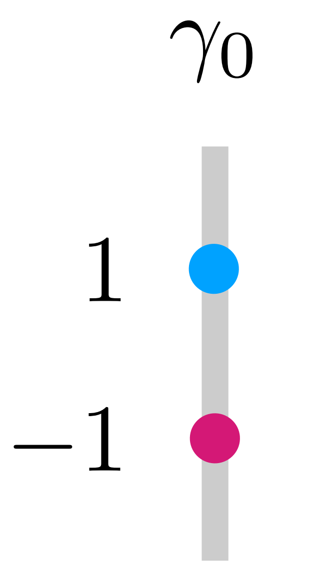
$$\begin{bmatrix} \boxed{\begin{matrix} \gamma_{00} \\ \gamma_{10} \\ \gamma_{20} \\ \gamma_{30} \\ \gamma_{40} \\ \vdots \\ \gamma_{T0} \end{matrix}} & \begin{matrix} \cdots \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ \\ \cdots \end{matrix} & \begin{matrix} \gamma_{0m'} \\ \gamma_{1m'} \\ \gamma_{2m'} \\ \gamma_{3m'} \\ \gamma_{4m'} \\ \vdots \\ \gamma_{Tm'} \end{matrix} & \begin{matrix} y_{00} \\ y_{10} \\ y_{20} \\ y_{30} \\ y_{40} \\ \vdots \\ y_{T0} \end{matrix} & \begin{matrix} \cdots \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ \\ \cdots \end{matrix} & \begin{matrix} y_{0m} \\ y_{1m} \\ y_{2m} \\ y_{3m} \\ y_{4m} \\ \vdots \\ y_{Tm} \end{matrix} \end{bmatrix} \Leftarrow \boxed{f \left(\begin{matrix} x_{00} & \cdots & x_{0n} \\ x_{10} & \cdots & x_{1n} \end{matrix} \right)} \begin{bmatrix} x_{20} & \cdots & x_{2n} \\ x_{30} & \cdots & x_{3n} \\ x_{40} & \cdots & x_{4n} \\ \vdots & & \vdots \\ x_{T0} & \cdots & x_{Tn} \end{bmatrix} \begin{bmatrix} \xi_{00} & \cdots & \xi_{0n'} \\ \xi_{10} & \cdots & \xi_{1n'} \\ \xi_{20} & \cdots & \xi_{2n'} \\ \xi_{30} & \cdots & \xi_{3n'} \\ \xi_{40} & \cdots & \xi_{4n'} \\ \vdots & & \vdots \\ \xi_{T0} & \cdots & \xi_{Tn'} \end{bmatrix}$$

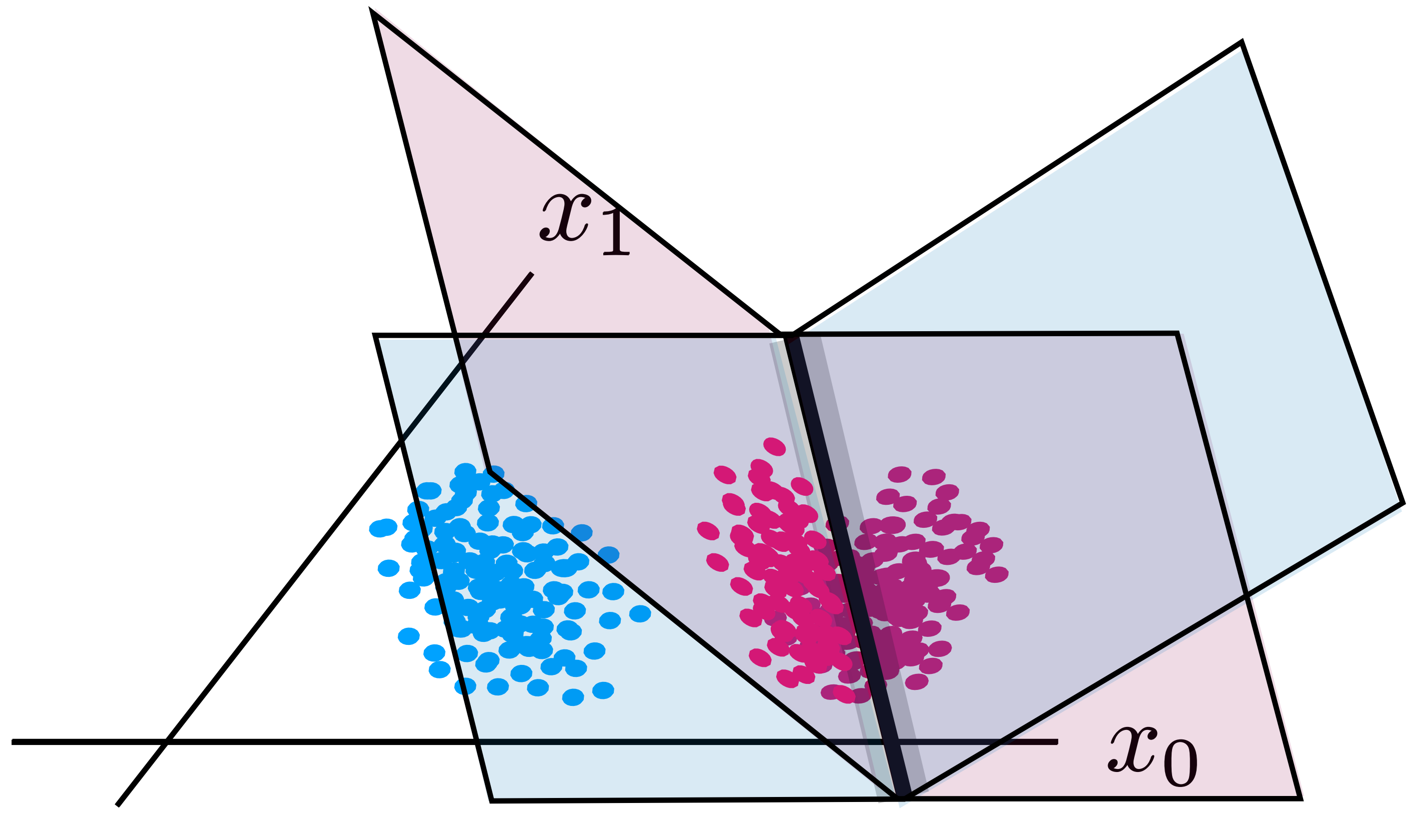
“Binary classifier”

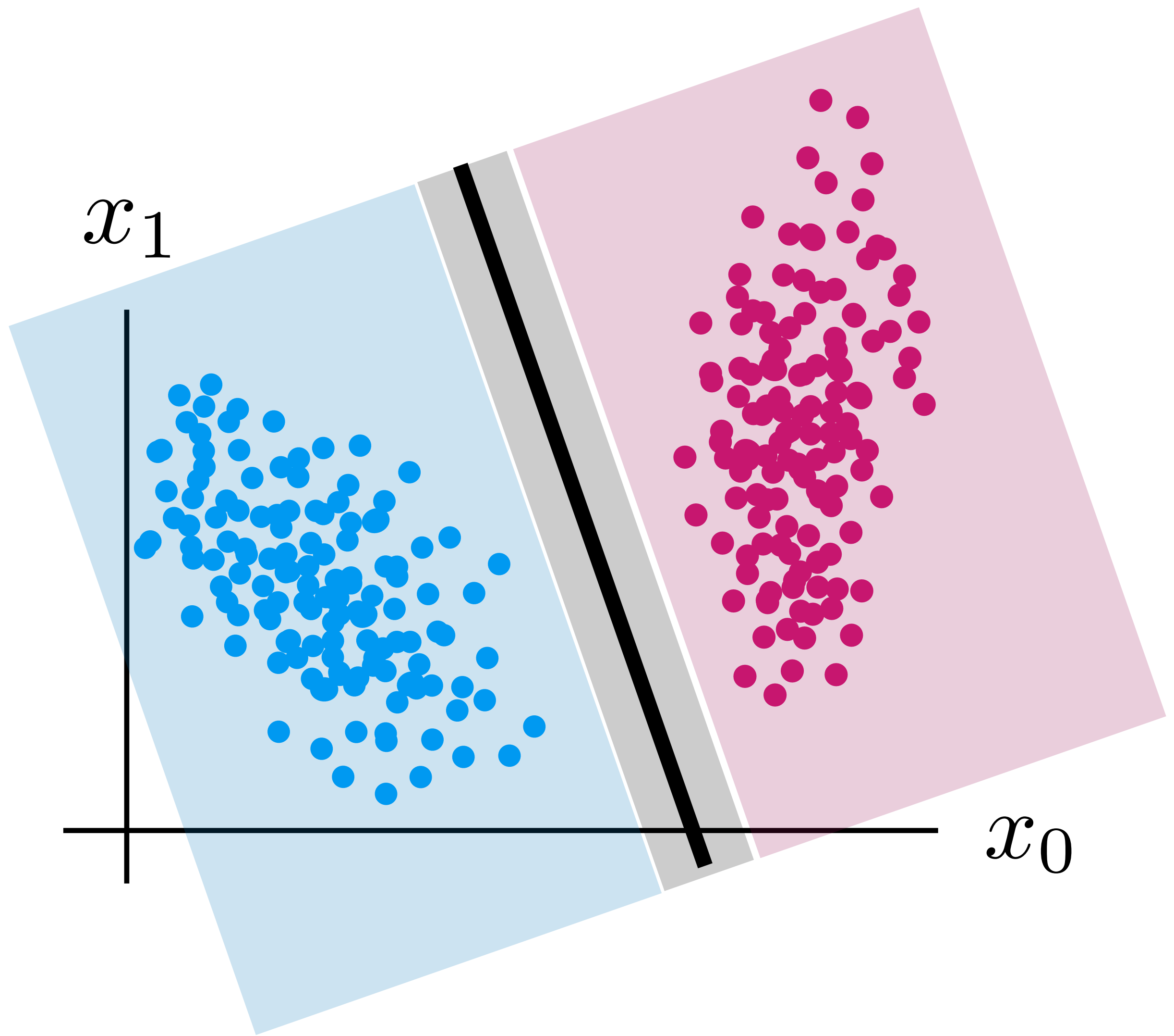
Support Vector Machine (SVM)

COST: $$\min_{\theta} \quad \lambda ||\theta_{1:n}||_2^2 + \frac{1}{T} \sum_t \max\{0, 1 - \gamma_t(\theta_{1:n}^T x_t - \theta_0)\}$$

Soft boundary







Classification

OUTPUTS

(Dependent Variables)

$$\gamma_t = \mathrm{sgn}(\theta^T x_t - \theta_0)$$

INPUTS

(Independent Variables)

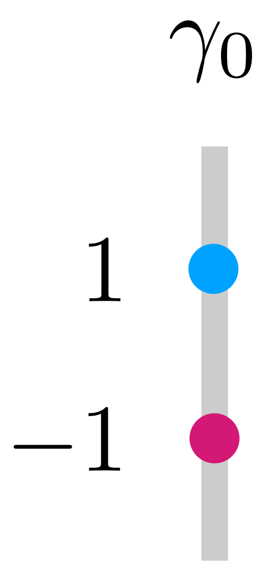
$$\begin{bmatrix} \boxed{\begin{matrix}\gamma_{00} \\ \gamma_{10} \\ \gamma_{20} \\ \gamma_{30} \\ \gamma_{40} \\ \vdots \\ \gamma_{T0}\end{matrix}} & \begin{matrix}\cdots \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ \\ \cdots\end{matrix} & \begin{matrix}\gamma_{0m'} \\ \gamma_{1m'} \\ \gamma_{2m'} \\ \gamma_{3m'} \\ \gamma_{4m'} \\ \vdots \\ \gamma_{Tm'}\end{matrix} & \begin{matrix}y_{00} \\ y_{10} \\ y_{20} \\ y_{30} \\ y_{40} \\ \vdots \\ y_{T0}\end{matrix} & \begin{matrix}\cdots \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ \\ \cdots\end{matrix} & \begin{matrix}y_{0m} \\ y_{1m} \\ y_{2m} \\ y_{3m} \\ y_{4m} \\ \vdots \\ y_{Tm}\end{matrix} \end{bmatrix} \Leftarrow \begin{bmatrix} \boxed{f\left(\begin{matrix} x_{00} & \cdots & x_{0n} \\ x_{10} & \cdots & x_{1n}\end{matrix}\right)} \\ \begin{matrix} x_{20} & \cdots & x_{2n} \\ x_{30} & \cdots & x_{3n} \\ x_{40} & \cdots & x_{4n} \\ \vdots & & \vdots \\ x_{T0} & \cdots & x_{Tn} \end{matrix} \end{bmatrix} \begin{bmatrix} \xi_{00} & \cdots & \xi_{0n'} \\ \xi_{10} & \cdots & \xi_{1n'} \\ \xi_{20} & \cdots & \xi_{2n'} \\ \xi_{30} & \cdots & \xi_{3n'} \\ \xi_{40} & \cdots & \xi_{4n'} \\ \vdots & & \vdots \\ \xi_{T0} & \cdots & \xi_{Tn'} \end{bmatrix}$$

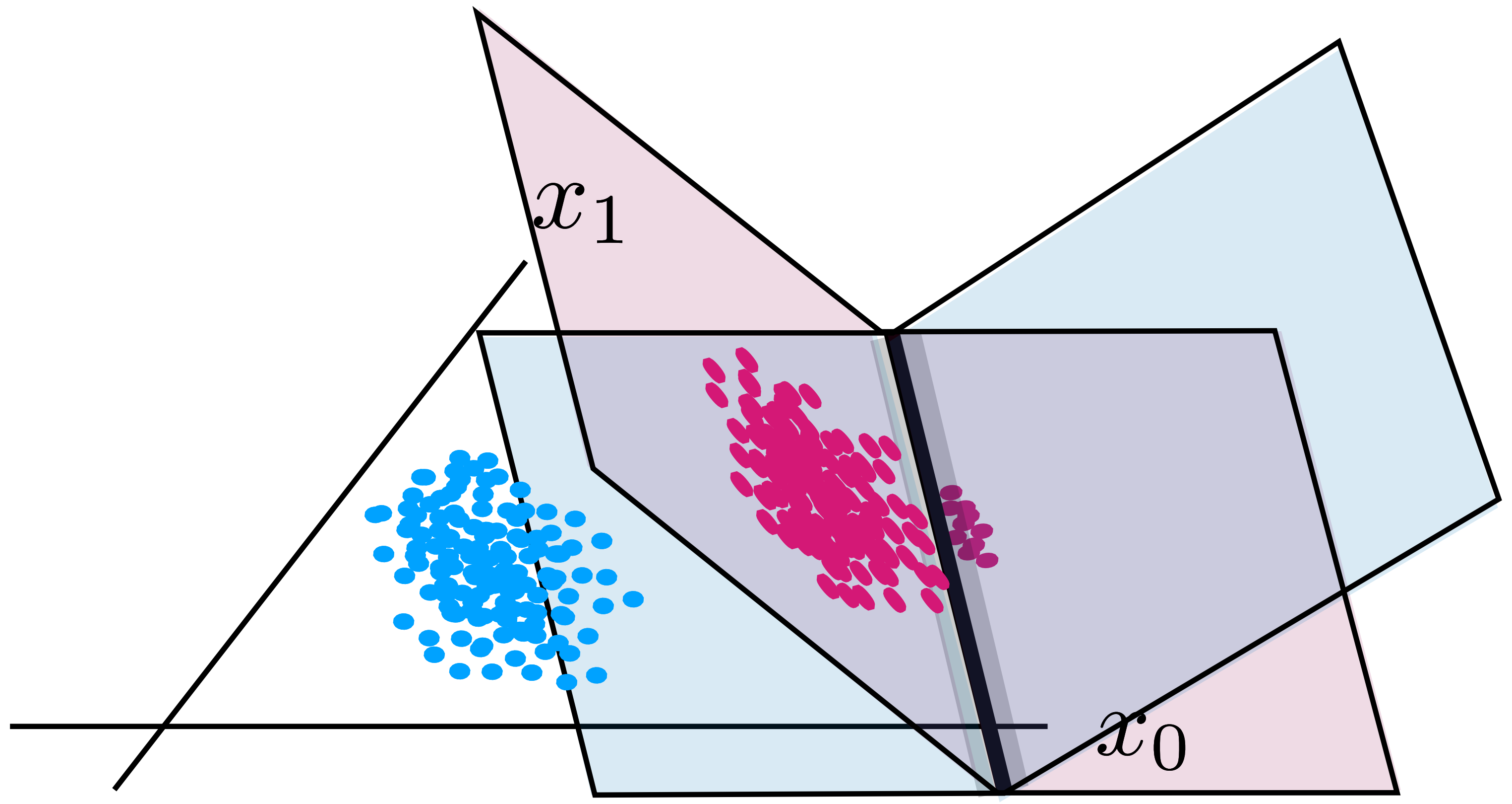
"Binary classifier"

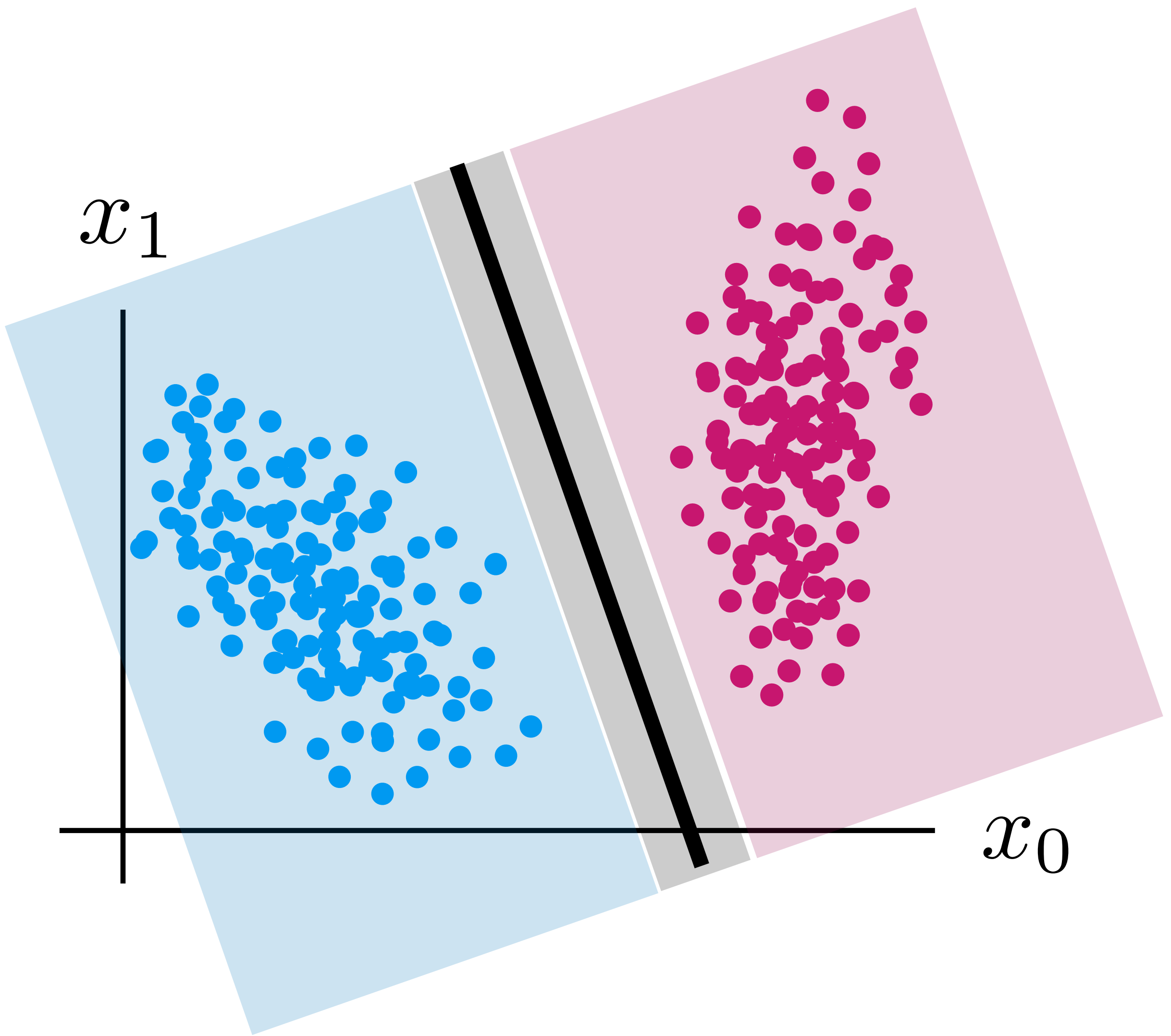
Support Vector Machine (SVM)

COST: $$\min_{\theta} \; \lambda||\theta_{1:n}||_2^2 + \frac{1}{T}\sum_t \max\{0, 1 - \gamma_t(\theta_{1:n}^T x_t - \theta_0)\}$$

Soft boundary







Classification

OUTPUTS

(Dependent Variables)

$$\gamma_t = \mathrm{sgn}(\theta^T x_t - \theta_0)$$

INPUTS

(Independent Variables)

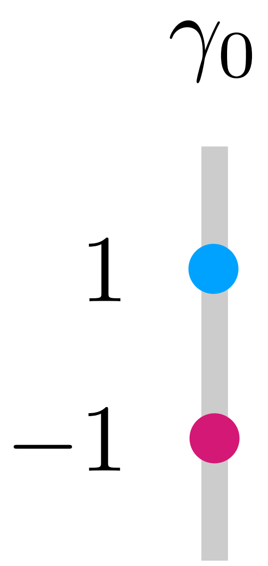
$$\begin{bmatrix} \gamma_{00} & \cdots & \gamma_{0m'} & y_{00} & \cdots & y_{0m} \\ \gamma_{10} & \cdots & \gamma_{1m'} & y_{10} & \cdots & y_{1m} \\ \gamma_{20} & \cdots & \gamma_{2m'} & y_{20} & \cdots & y_{2m} \\ \gamma_{30} & \cdots & \gamma_{3m'} & y_{30} & \cdots & y_{3m} \\ \gamma_{40} & \cdots & \gamma_{4m'} & y_{40} & \cdots & y_{4m} \\ \vdots & & \vdots & \vdots & & \vdots \\ \gamma_{T0} & \cdots & \gamma_{Tm'} & y_{T0} & \cdots & y_{Tm} \end{bmatrix} \Leftarrow f\left[\left(\begin{matrix} x_{00} & \cdots & x_{0n} \\ x_{10} & \cdots & x_{1n} \end{matrix}\right) \atop \begin{matrix} x_{20} & \cdots & x_{2n} \\ x_{30} & \cdots & x_{3n} \\ x_{40} & \cdots & x_{4n} \\ \vdots & & \vdots \\ x_{T0} & \cdots & x_{Tn} \end{matrix}\right] \begin{bmatrix} \xi_{00} & \cdots & \xi_{0n'} \\ \xi_{10} & \cdots & \xi_{1n'} \\ \xi_{20} & \cdots & \xi_{2n'} \\ \xi_{30} & \cdots & \xi_{3n'} \\ \xi_{40} & \cdots & \xi_{4n'} \\ \vdots & & \vdots \\ \xi_{T0} & \cdots & \xi_{Tn'} \end{bmatrix}$$

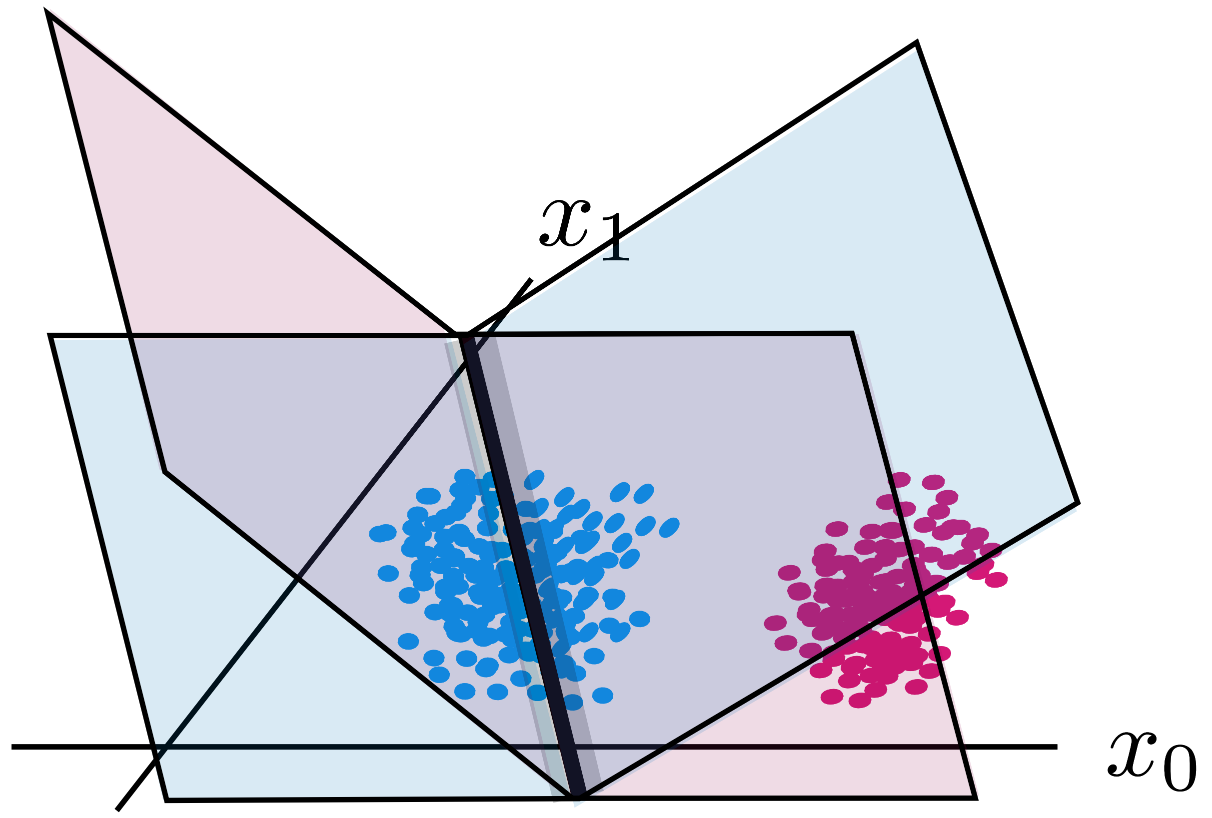
“Binary classifier”

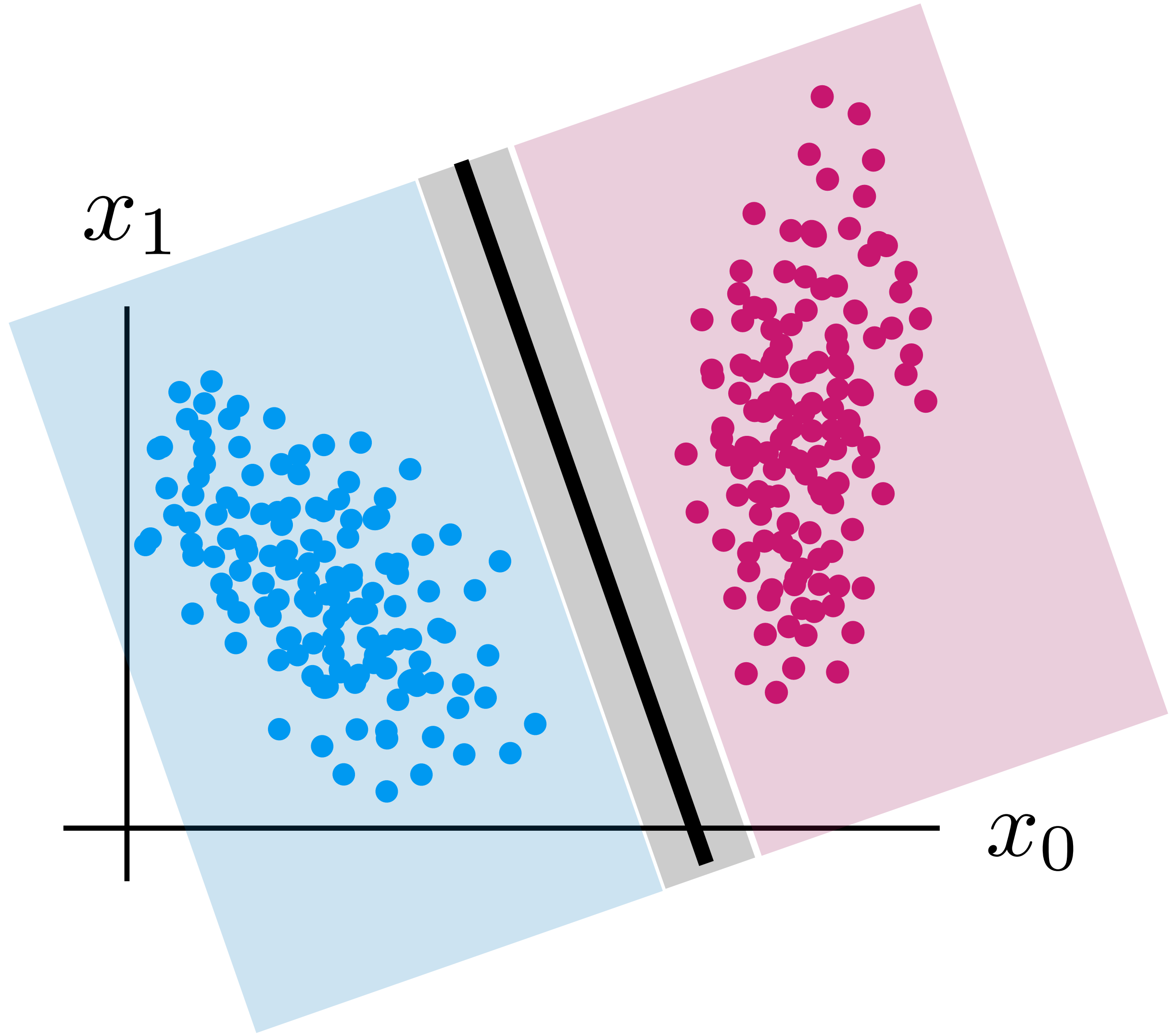
Support Vector Machine (SVM)

COST: $$\min_{\theta} \; \lambda||\theta_{1:n}||_2^2 + \frac{1}{T}\sum_t \max\{0, 1 - \gamma_t(\theta_{1:n}^T x_t - \theta_0)\}$$

Soft boundary







Classification

OUTPUTS

(Dependent Variables)

$$\gamma_t = \mathrm{sgn}(\theta^T x_t - \theta_0)$$

INPUTS

(Independent Variables)

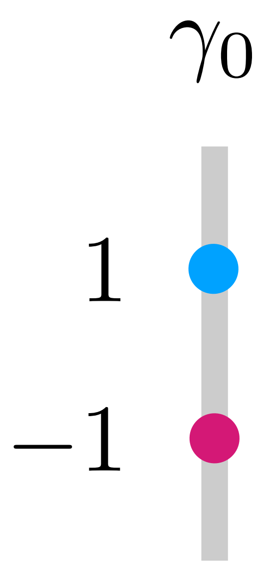
$$\begin{bmatrix} \boxed{\begin{matrix} \gamma_{00} \\ \gamma_{10} \\ \gamma_{20} \\ \gamma_{30} \\ \gamma_{40} \\ \vdots \\ \gamma_{T0} \end{matrix}} & \begin{matrix} \cdots \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ \\ \cdots \end{matrix} & \begin{matrix} \gamma_{0m'} \\ \gamma_{1m'} \\ \gamma_{2m'} \\ \gamma_{3m'} \\ \gamma_{4m'} \\ \vdots \\ \gamma_{Tm'} \end{matrix} & \begin{matrix} y_{00} \\ y_{10} \\ y_{20} \\ y_{30} \\ y_{40} \\ \vdots \\ y_{T0} \end{matrix} & \begin{matrix} \cdots \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ \\ \cdots \end{matrix} & \begin{matrix} y_{0m} \\ y_{1m} \\ y_{2m} \\ y_{3m} \\ y_{4m} \\ \vdots \\ y_{Tm} \end{matrix} \end{bmatrix} \Leftarrow \boxed{f \left[\left(\begin{matrix} x_{00} & \cdots & x_{0n} \\ x_{10} & \cdots & x_{1n} \end{matrix} \right) \right.} \begin{matrix} \\ \\ x_{20} & \cdots & x_{2n} \\ x_{30} & \cdots & x_{3n} \\ x_{40} & \cdots & x_{4n} \\ \vdots & & \vdots \\ x_{T0} & \cdots & x_{Tn} \end{matrix} \Bigg] \begin{bmatrix} \xi_{00} & \cdots & \xi_{0n'} \\ \xi_{10} & \cdots & \xi_{1n'} \\ \xi_{20} & \cdots & \xi_{2n'} \\ \xi_{30} & \cdots & \xi_{3n'} \\ \xi_{40} & \cdots & \xi_{4n'} \\ \vdots & & \vdots \\ \xi_{T0} & \cdots & \xi_{Tn'} \end{bmatrix}$$

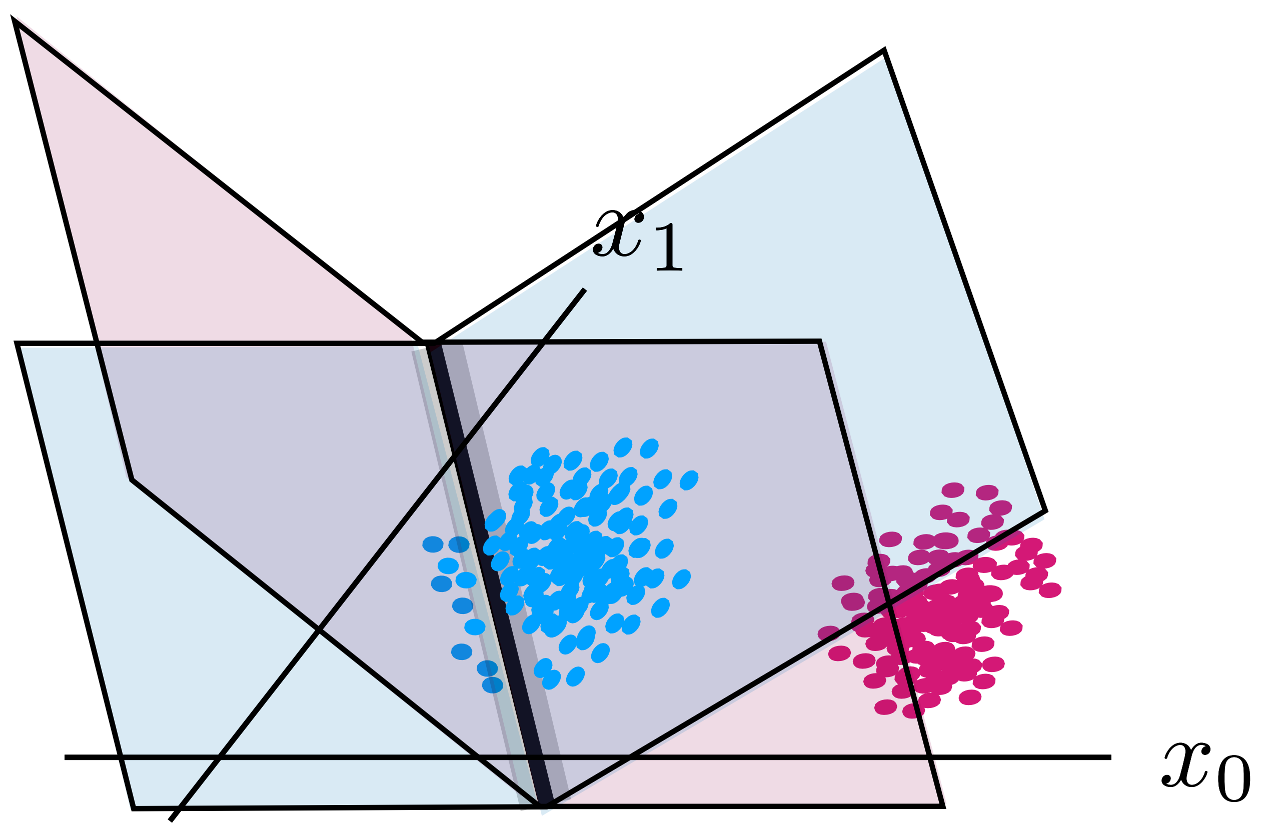
"Binary classifier"

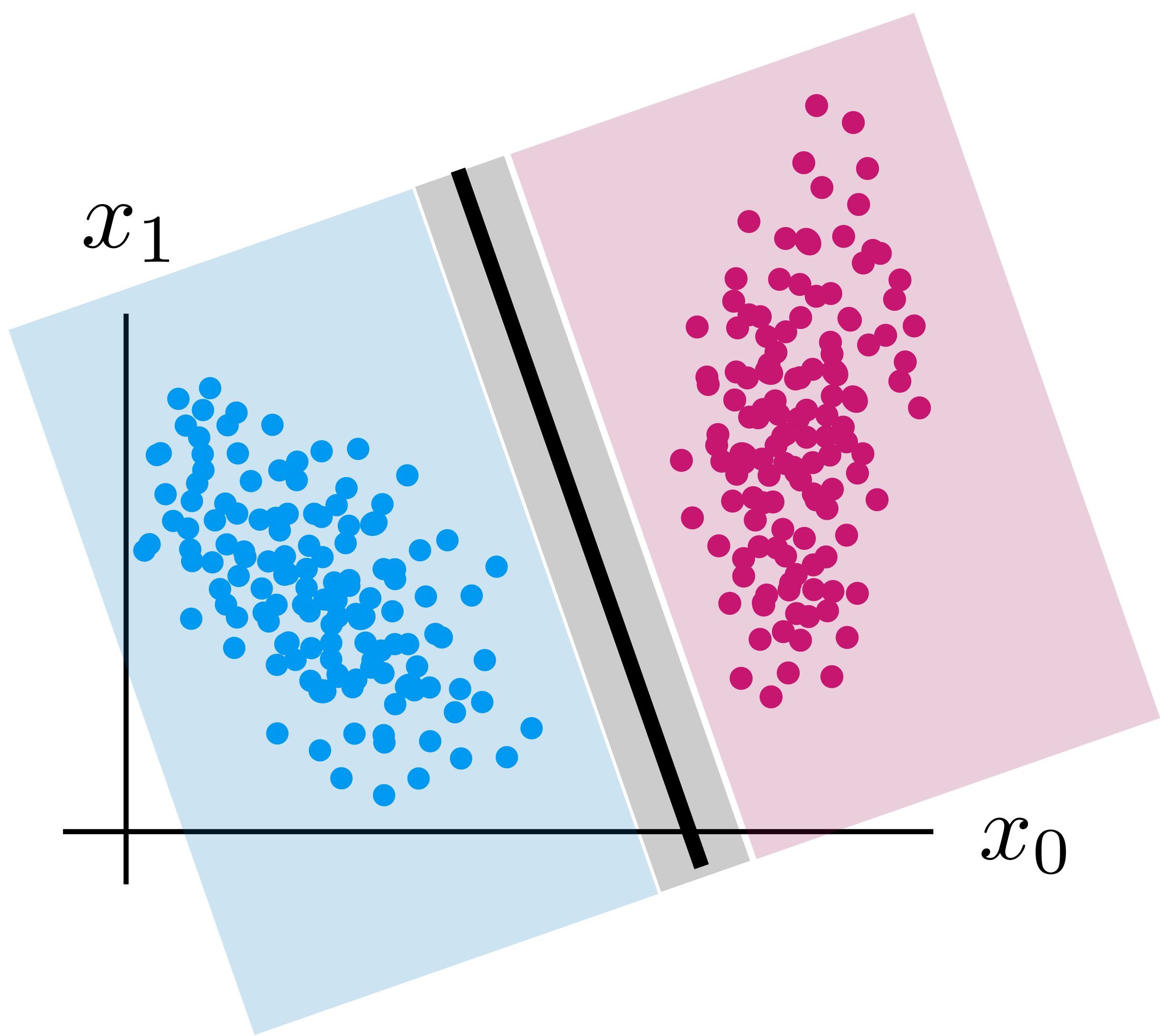
Support Vector Machine (SVM)

COST: $$\min_{\theta} \; \lambda ||\theta_{1:n}||_2^2 + \frac{1}{T} \sum_t \max\{0, 1 - \gamma_t(\theta_{1:n}^T x_t - \theta_0)\}$$

Soft boundary







Classification

OUTPUTS

(Dependent Variables)

$$\gamma_t = \mathrm{sgn}(\theta^T x_t - \theta_0)$$

INPUTS

(Independent Variables)

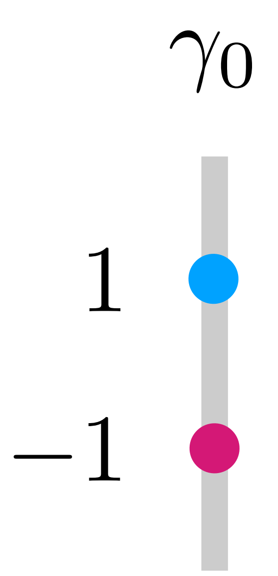
$$\begin{bmatrix} \gamma_{00} & \cdots & \gamma_{0m'} & y_{00} & \cdots & y_{0m} \\ \gamma_{10} & \cdots & \gamma_{1m'} & y_{10} & \cdots & y_{1m} \\ \gamma_{20} & \cdots & \gamma_{2m'} & y_{20} & \cdots & y_{2m} \\ \gamma_{30} & \cdots & \gamma_{3m'} & y_{30} & \cdots & y_{3m} \\ \gamma_{40} & \cdots & \gamma_{4m'} & y_{40} & \cdots & y_{4m} \\ \vdots & & \vdots & \vdots & & \vdots \\ \gamma_{T0} & \cdots & \gamma_{Tm'} & y_{T0} & \cdots & y_{Tm} \end{bmatrix} \Leftarrow f\begin{bmatrix} (x_{00} & \cdots & x_{0n}) \\ x_{10} & \cdots & x_{1n} \\ x_{20} & \cdots & x_{2n} \\ x_{30} & \cdots & x_{3n} \\ x_{40} & \cdots & x_{4n} \\ \vdots & & \vdots \\ x_{T0} & \cdots & x_{Tn} \end{bmatrix} \begin{bmatrix} \xi_{00} & \cdots & \xi_{0n'} \\ \xi_{10} & \cdots & \xi_{1n'} \\ \xi_{20} & \cdots & \xi_{2n'} \\ \xi_{30} & \cdots & \xi_{3n'} \\ \xi_{40} & \cdots & \xi_{4n'} \\ \vdots & & \vdots \\ \xi_{T0} & \cdots & \xi_{Tn'} \end{bmatrix}$$

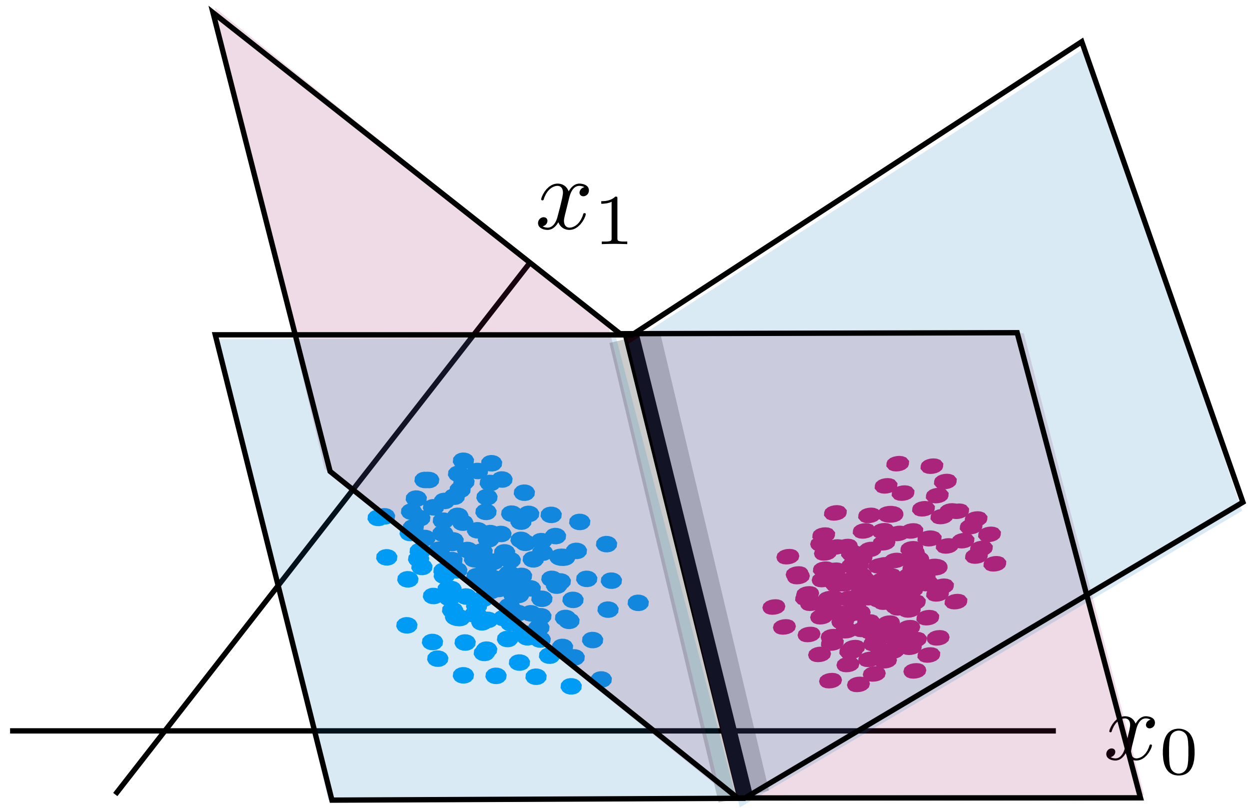
"Binary classifier"

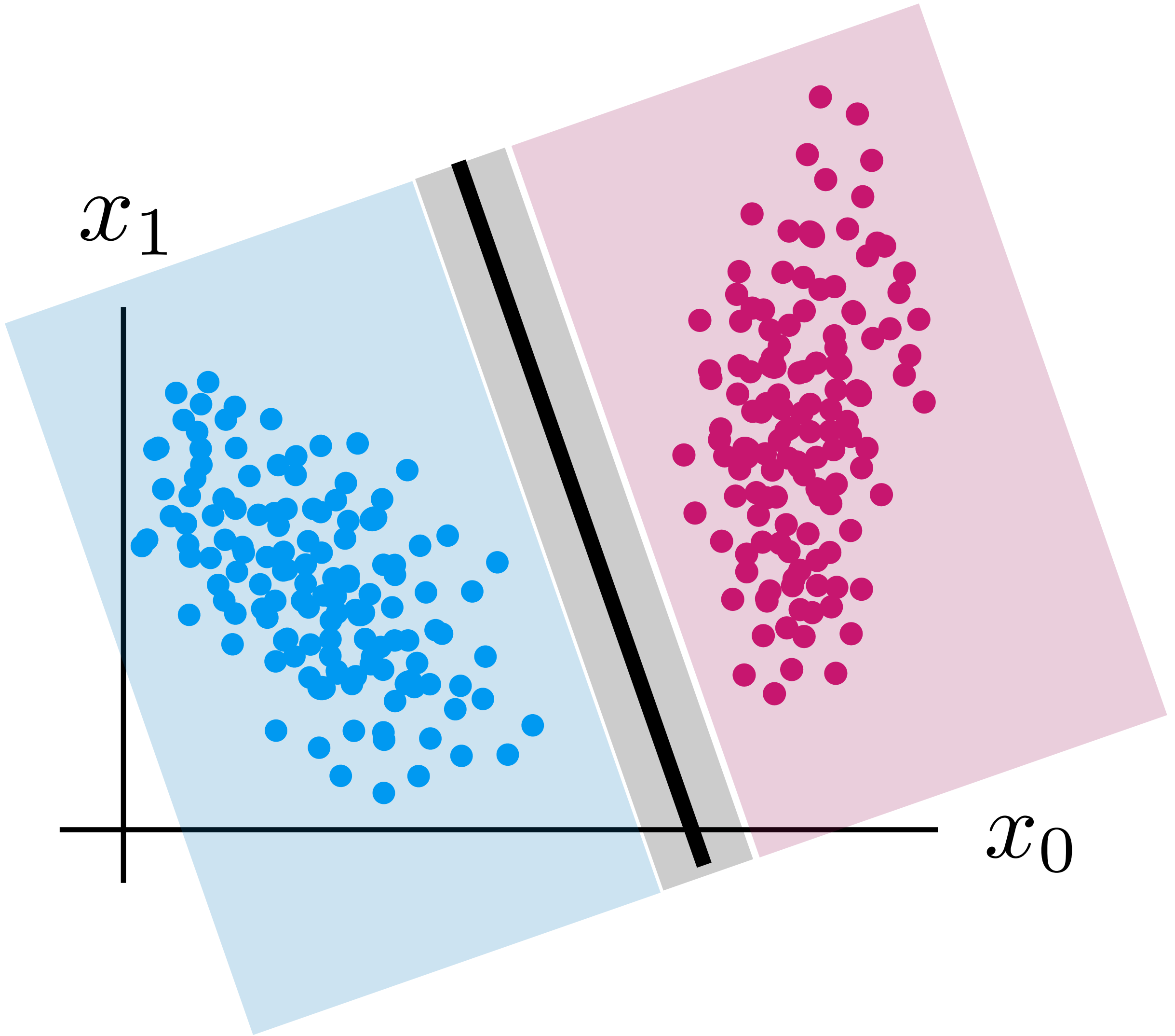
Support Vector Machine (SVM)

COST: $$\min_{\theta} \; \lambda||\theta_{1:n}||_2^2 + \frac{1}{T}\sum_t \max\{0, 1 - \gamma_t(\theta_{1:n}^T x_t - \theta_0)\}$$

Soft boundary







Classification

OUTPUTS

(Dependent Variables)

$$\gamma_t = \mathrm{sgn}(\theta^T x_t - \theta_0)$$

INPUTS

(Independent Variables)

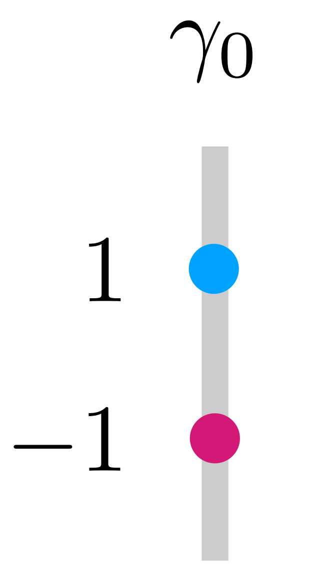
$$\begin{bmatrix} \boxed{\begin{matrix} \gamma_{00} \\ \gamma_{10} \\ \gamma_{20} \\ \gamma_{30} \\ \gamma_{40} \\ \vdots \\ \gamma_{T0} \end{matrix}} & \begin{matrix} \cdots \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ \\ \cdots \end{matrix} & \begin{matrix} \gamma_{0m'} \\ \gamma_{1m'} \\ \gamma_{2m'} \\ \gamma_{3m'} \\ \gamma_{4m'} \\ \vdots \\ \gamma_{Tm'} \end{matrix} & \begin{matrix} y_{00} \\ y_{10} \\ y_{20} \\ y_{30} \\ y_{40} \\ \vdots \\ y_{T0} \end{matrix} & \begin{matrix} \cdots \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ \\ \cdots \end{matrix} & \begin{matrix} y_{0m} \\ y_{1m} \\ y_{2m} \\ y_{3m} \\ y_{4m} \\ \vdots \\ y_{Tm} \end{matrix} \end{bmatrix} \Leftarrow \boxed{f \Big[\Big(\begin{matrix} x_{00} & \cdots & x_{0n} \\ x_{10} & \cdots & x_{1n} \end{matrix} \Big)} \begin{matrix} \\ x_{20} & \cdots & x_{2n} \\ x_{30} & \cdots & x_{3n} \\ x_{40} & \cdots & x_{4n} \\ \vdots & & \vdots \\ x_{T0} & \cdots & x_{Tn} \end{matrix} \Big] \begin{bmatrix} \xi_{00} & \cdots & \xi_{0n'} \\ \xi_{10} & \cdots & \xi_{1n'} \\ \xi_{20} & \cdots & \xi_{2n'} \\ \xi_{30} & \cdots & \xi_{3n'} \\ \xi_{40} & \cdots & \xi_{4n'} \\ \vdots & & \vdots \\ \xi_{T0} & \cdots & \xi_{Tn'} \end{bmatrix}$$

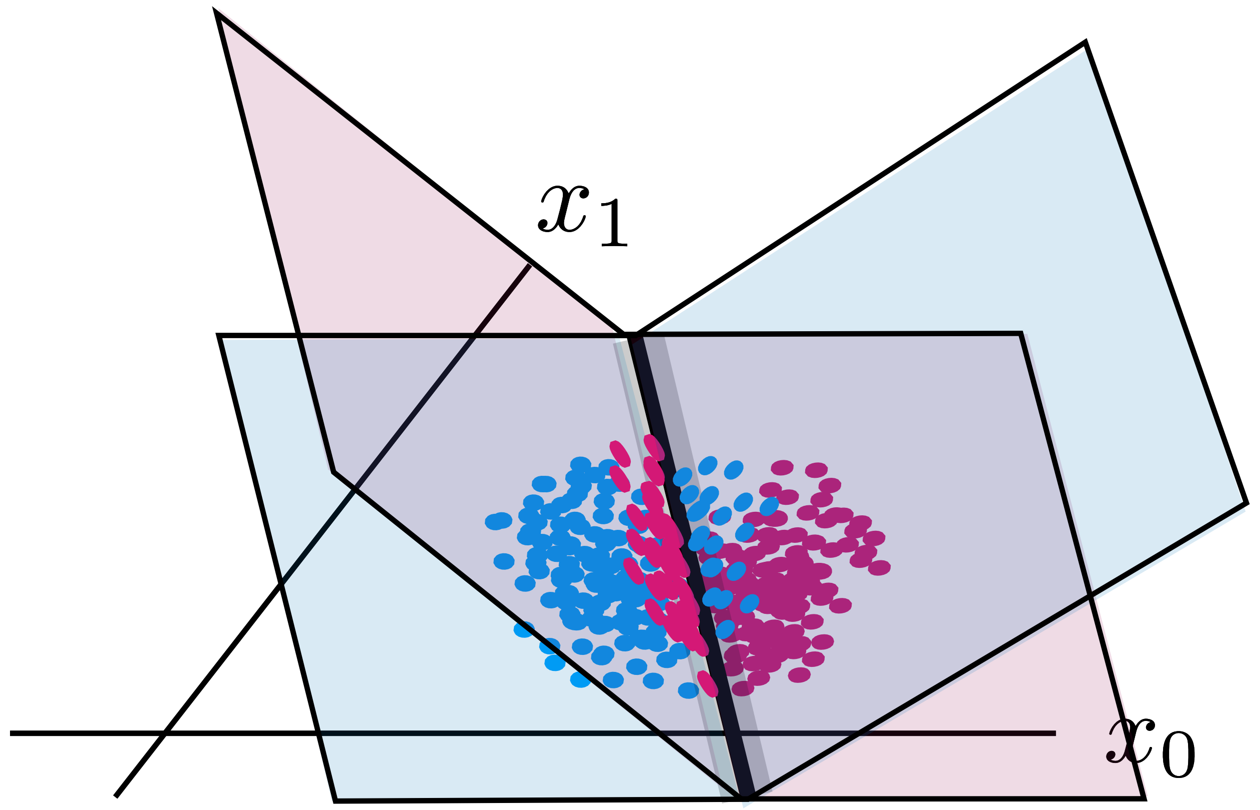
"Binary classifier"

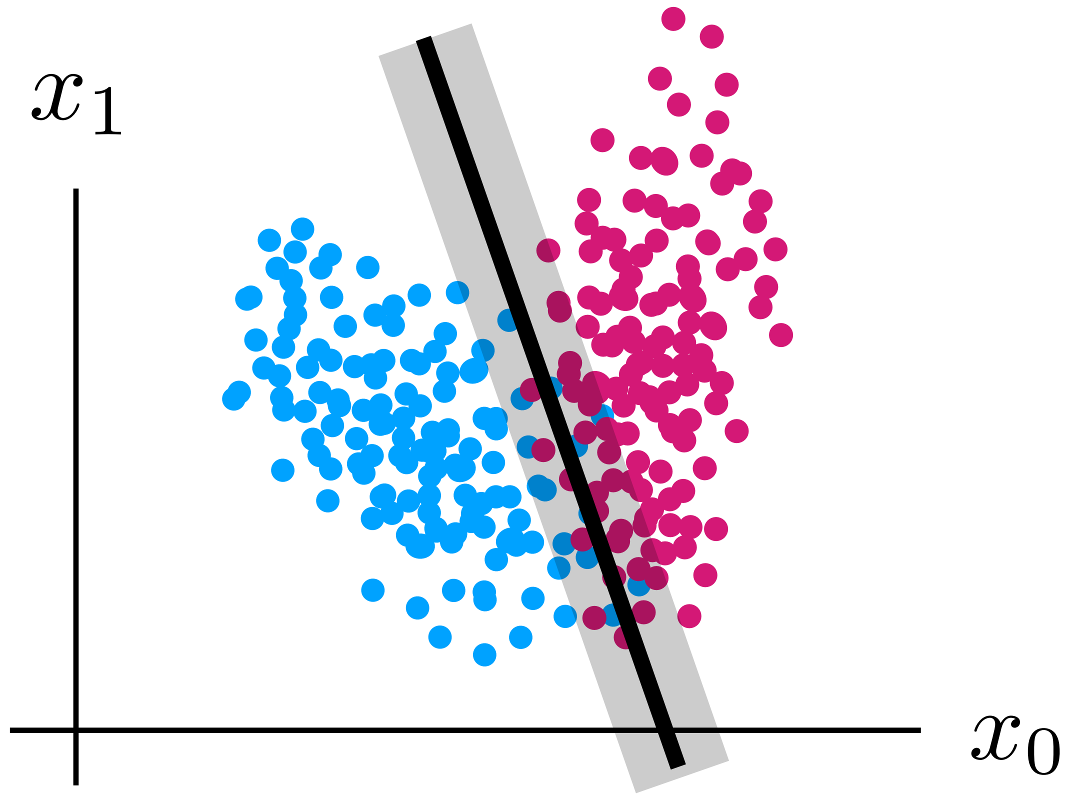
Support Vector Machine (SVM)

COST: $$\min_{\theta} \quad \lambda||\theta_{1:n}||_2^2 + \frac{1}{T}\sum_t \max\{0, 1 - \gamma_t(\theta_{1:n}^T x_t - \theta_0)\}$$

Soft boundary







Classification

OUTPUTS

(Dependent Variables)

$$\gamma_t = \text{sgn}(\theta^T x_t - \theta_0)$$

INPUTS

(Independent Variables)

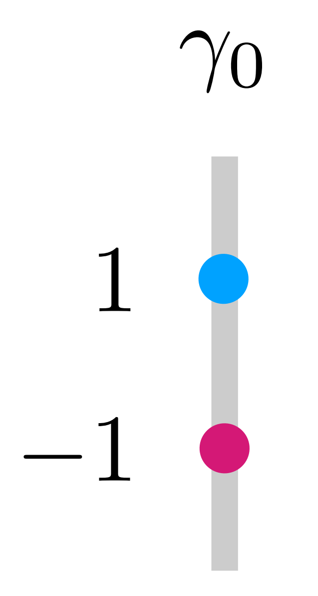
$$\begin{bmatrix} \gamma_{00} & \cdots & \gamma_{0m'} & y_{00} & \cdots & y_{0m} \\ \gamma_{10} & \cdots & \gamma_{1m'} & y_{10} & \cdots & y_{1m} \\ \gamma_{20} & \cdots & \gamma_{2m'} & y_{20} & \cdots & y_{2m} \\ \gamma_{30} & \cdots & \gamma_{3m'} & y_{30} & \cdots & y_{3m} \\ \gamma_{40} & \cdots & \gamma_{4m'} & y_{40} & \cdots & y_{4m} \\ \vdots & & \vdots & \vdots & & \vdots \\ \gamma_{T0} & \cdots & \gamma_{Tm'} & y_{T0} & \cdots & y_{Tm} \end{bmatrix} \Leftarrow f \begin{bmatrix} \left(x_{00} \right. & \cdots & \left. x_{0n} \right) \\ x_{10} & \cdots & x_{1n} \\ x_{20} & \cdots & x_{2n} \\ x_{30} & \cdots & x_{3n} \\ x_{40} & \cdots & x_{4n} \\ \vdots & & \vdots \\ x_{T0} & \cdots & x_{Tn} \end{bmatrix} \begin{bmatrix} \xi_{00} & \cdots & \xi_{0n'} \\ \xi_{10} & \cdots & \xi_{1n'} \\ \xi_{20} & \cdots & \xi_{2n'} \\ \xi_{30} & \cdots & \xi_{3n'} \\ \xi_{40} & \cdots & \xi_{4n'} \\ \vdots & & \vdots \\ \xi_{T0} & \cdots & \xi_{Tn'} \end{bmatrix}$$

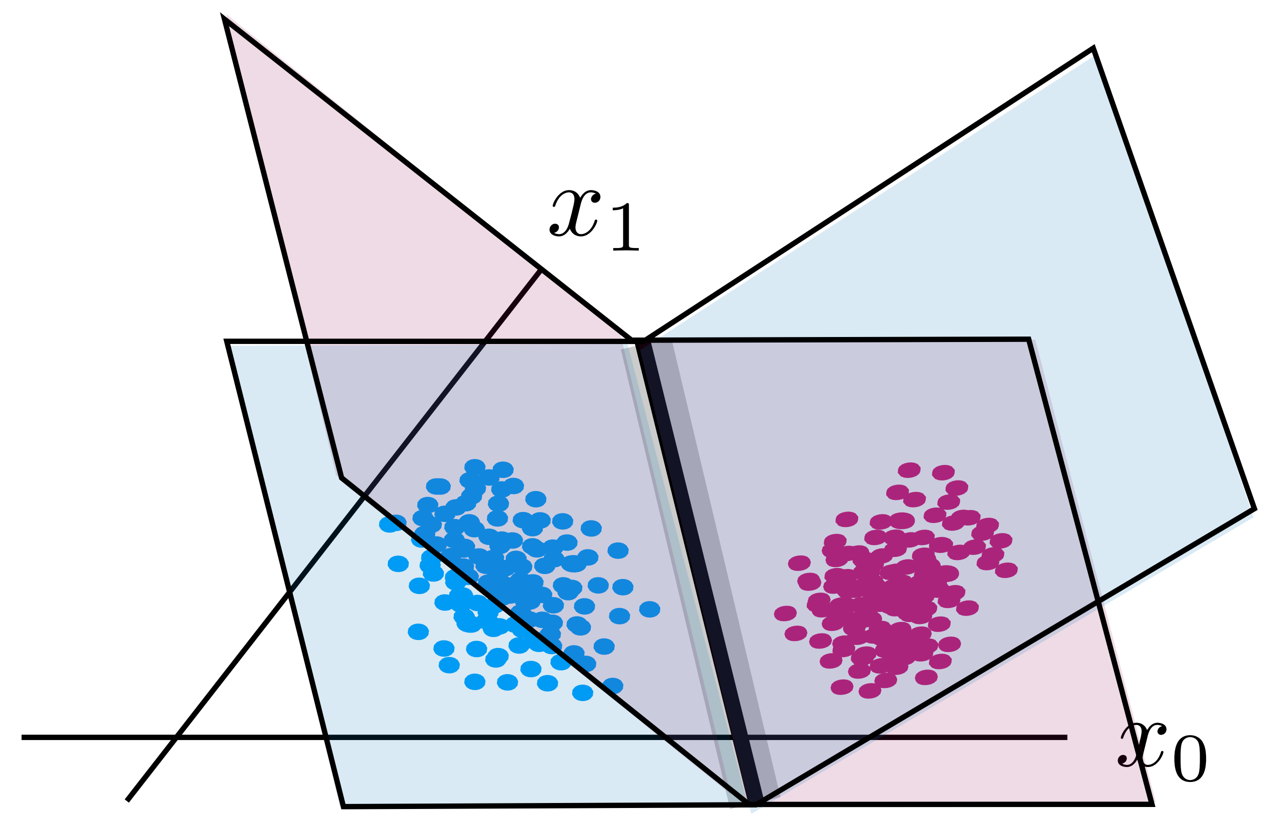
"Binary classifier"

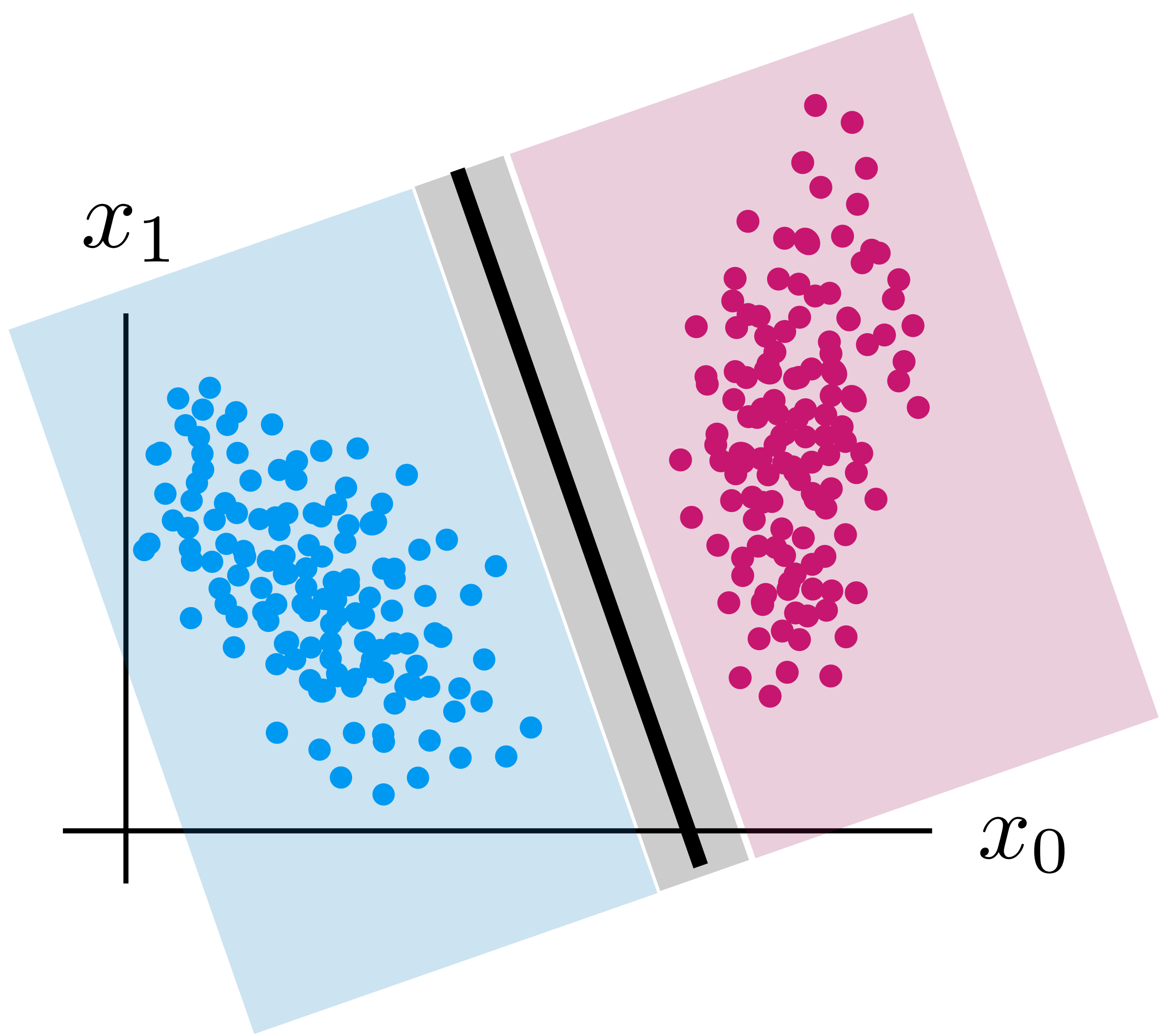
Support Vector Machine (SVM)

COST: $$\min_{\theta} \; \lambda ||\theta_{1:n}||_2^2 + \frac{1}{T} \sum_t \max\{0, 1 - \gamma_t(\theta_{1:n}^T x_t - \theta_0)\}$$

Soft boundary







Classification

OUTPUTS

(Dependent Variables)

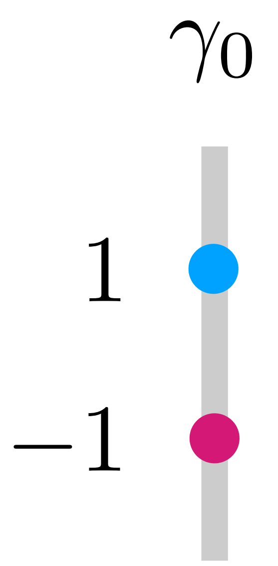
$$\gamma_t = \mathrm{sgn}(\theta^T x_t - \theta_0)$$

INPUTS

(Independent Variables)

$$\begin{bmatrix} \boxed{\begin{matrix}\gamma_{00} \\ \gamma_{10} \\ \gamma_{20} \\ \gamma_{30} \\ \gamma_{40} \\ \vdots \\ \gamma_{T0}\end{matrix}} & \begin{matrix}\cdots \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ \\ \cdots\end{matrix} & \begin{matrix}\gamma_{0m'} \\ \gamma_{1m'} \\ \gamma_{2m'} \\ \gamma_{3m'} \\ \gamma_{4m'} \\ \vdots \\ \gamma_{Tm'}\end{matrix} & \begin{matrix}y_{00} \\ y_{10} \\ y_{20} \\ y_{30} \\ y_{40} \\ \vdots \\ y_{T0}\end{matrix} & \begin{matrix}\cdots \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ \\ \cdots\end{matrix} & \begin{matrix}y_{0m} \\ y_{1m} \\ y_{2m} \\ y_{3m} \\ y_{4m} \\ \vdots \\ y_{Tm}\end{matrix} \end{bmatrix} \Leftarrow \boxed{f\left[\begin{pmatrix} x_{00} & \cdots & x_{0n} \\ x_{10} & \cdots & x_{1n} \end{pmatrix}\right.} \begin{matrix} \\ \\ x_{20} & \cdots & x_{2n} \\ x_{30} & \cdots & x_{3n} \\ x_{40} & \cdots & x_{4n} \\ \vdots & & \vdots \\ x_{T0} & \cdots & x_{Tn} \end{matrix} \Bigg] \begin{bmatrix} \xi_{00} & \cdots & \xi_{0n'} \\ \xi_{10} & \cdots & \xi_{1n'} \\ \xi_{20} & \cdots & \xi_{2n'} \\ \xi_{30} & \cdots & \xi_{3n'} \\ \xi_{40} & \cdots & \xi_{4n'} \\ \vdots & & \vdots \\ \xi_{T0} & \cdots & \xi_{Tn'} \end{bmatrix}$$

"Binary classifier"



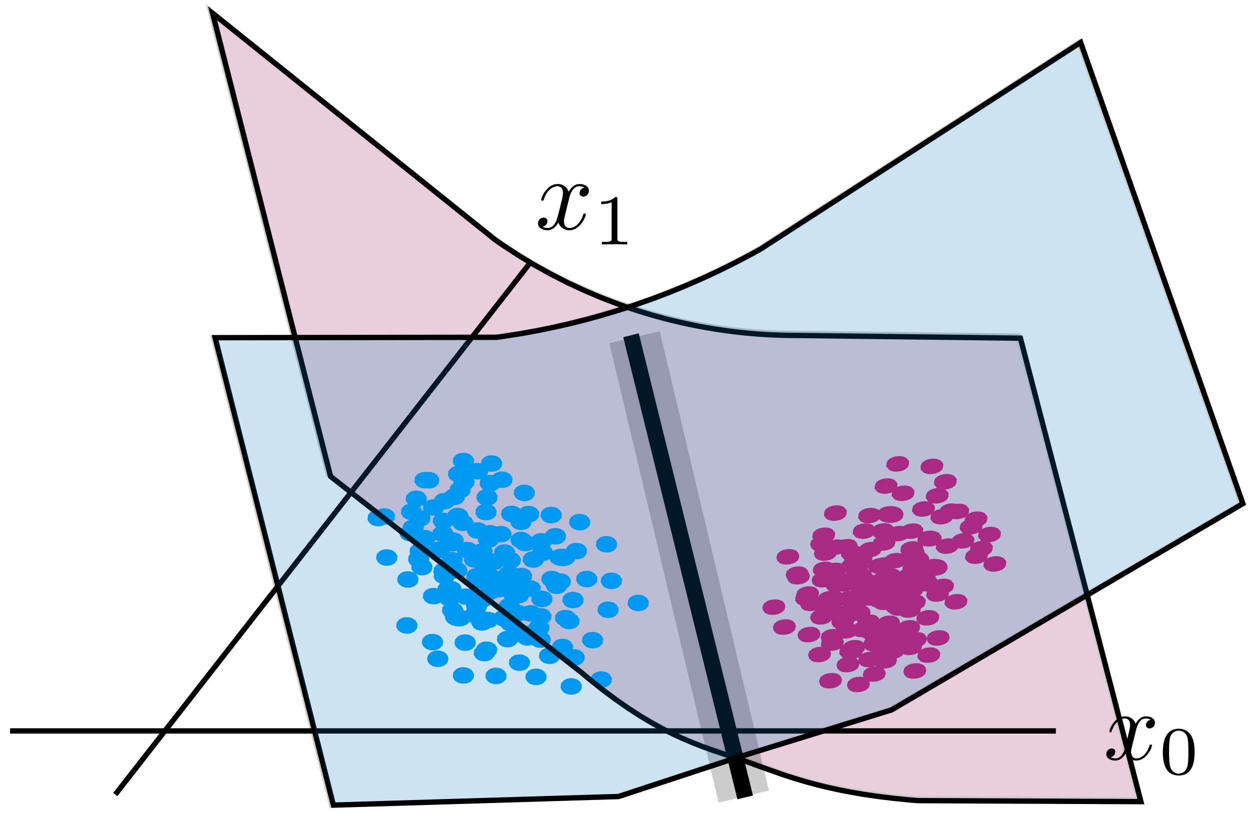
Support Vector Machine (SVM)

Logistic Regression (smoothed)

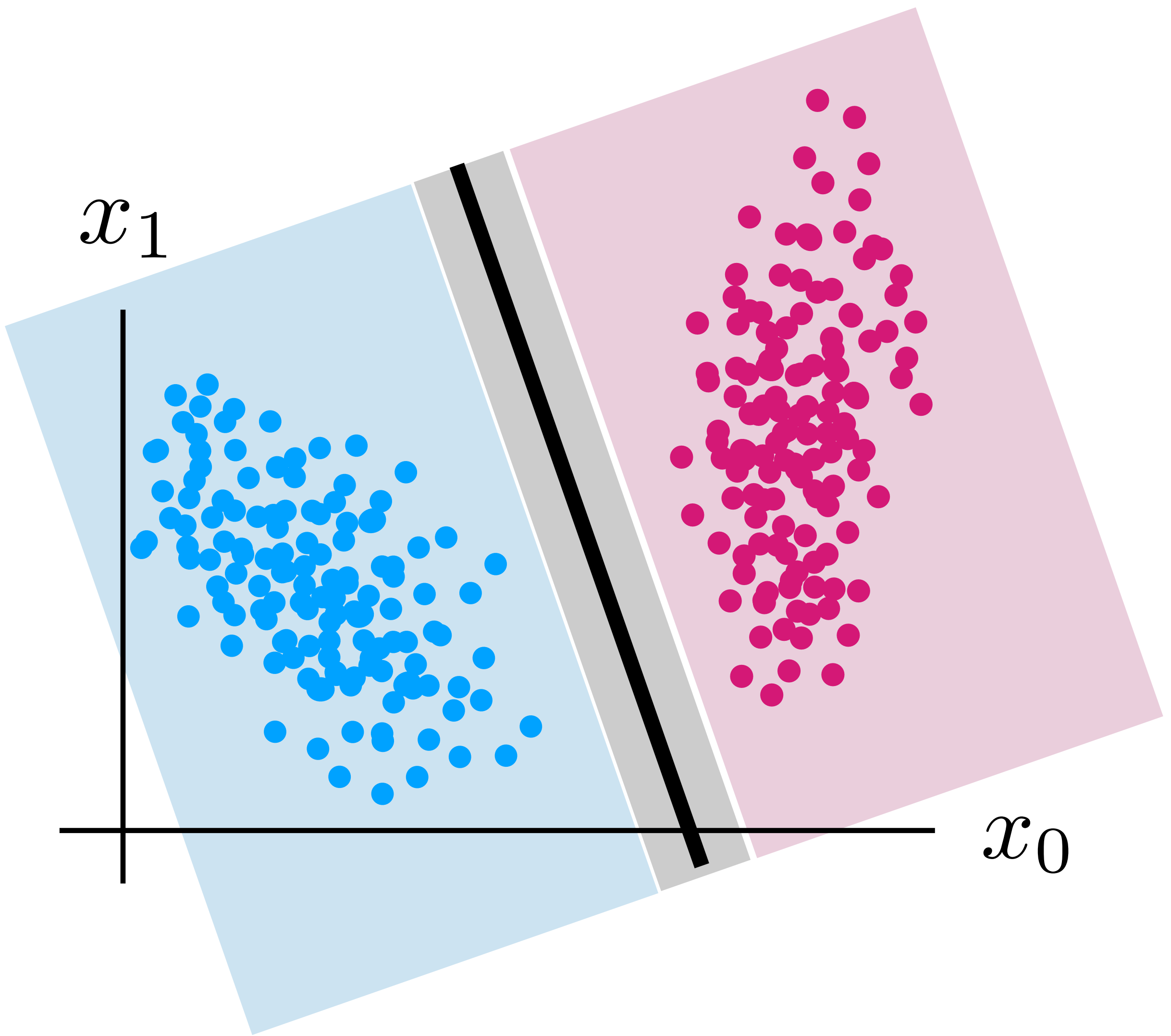
COST:

$$\min_{\theta} \quad \lambda||\theta_{1:n}||_2^2 + \frac{1}{T}\sum_t \max\{0, 1 - \gamma_t(\theta_{1:n}^T x_t - \theta_0)\}$$

$$\lambda||\theta_{1:n}||_2^2 + \frac{1}{T}\sum_t \log\left(e^{-\gamma_t\left(\theta_{1:n}^T x_t - \theta_0\right)+1}\right)$$



Soft boundary



Z

OUTPUTS (Dependent Variables)

$$\gamma_t = \mathrm{sgn}(\theta^T x_t - \theta_0)$$

INPUTS (Independent Variables)

$$\begin{bmatrix} \boxed{\begin{matrix}\gamma_{00}\\ \gamma_{10}\\ \gamma_{20}\\ \gamma_{30}\\ \gamma_{40}\\ \vdots \\ \gamma_{T0}\end{matrix}} & \begin{matrix}\cdots\\ \cdots\\ \cdots\\ \cdots\\ \cdots\\ \\ \cdots\end{matrix} & \begin{matrix}\gamma_{0m'}\\ \gamma_{1m'}\\ \gamma_{2m'}\\ \gamma_{3m'}\\ \gamma_{4m'}\\ \vdots \\ \gamma_{Tm'}\end{matrix} & \begin{matrix}y_{00}\\ y_{10}\\ y_{20}\\ y_{30}\\ y_{40}\\ \vdots \\ y_{T0}\end{matrix} & \begin{matrix}\cdots\\ \cdots\\ \cdots\\ \cdots\\ \cdots\\ \\ \cdots\end{matrix} & \begin{matrix}y_{0m}\\ y_{1m}\\ y_{2m}\\ y_{3m}\\ y_{4m}\\ \vdots \\ y_{Tm}\end{matrix} \end{bmatrix} \Leftarrow \begin{bmatrix} h_{00}(x_0) & \cdots & h_{0n}(x_0) \\ h_{10}(x_1) & \cdots & h_{1n}(x_1) \\ h_{20}(x_2) & \cdots & h_{2n}(x_2) \\ h_{30}(x_3) & \cdots & h_{3n}(x_3) \\ h_{40}(x_4) & \cdots & h_{4n}(x_4) \\ \vdots & & \vdots \\ h_{T0}(x_T) & \cdots & h_{Tn}(x_T) \end{bmatrix} \Leftarrow \begin{bmatrix} x_{00} & \cdots & x_{0n} \\ x_{10} & \cdots & x_{1n} \\ x_{20} & \cdots & x_{2n} \\ x_{30} & \cdots & x_{3n} \\ x_{40} & \cdots & x_{4n} \\ \vdots & & \vdots \\ x_{T0} & \cdots & x_{Tn} \end{bmatrix}$$

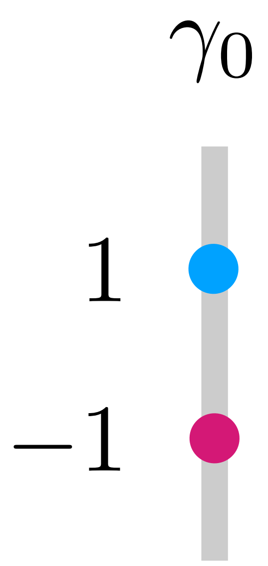
"Binary classifier"

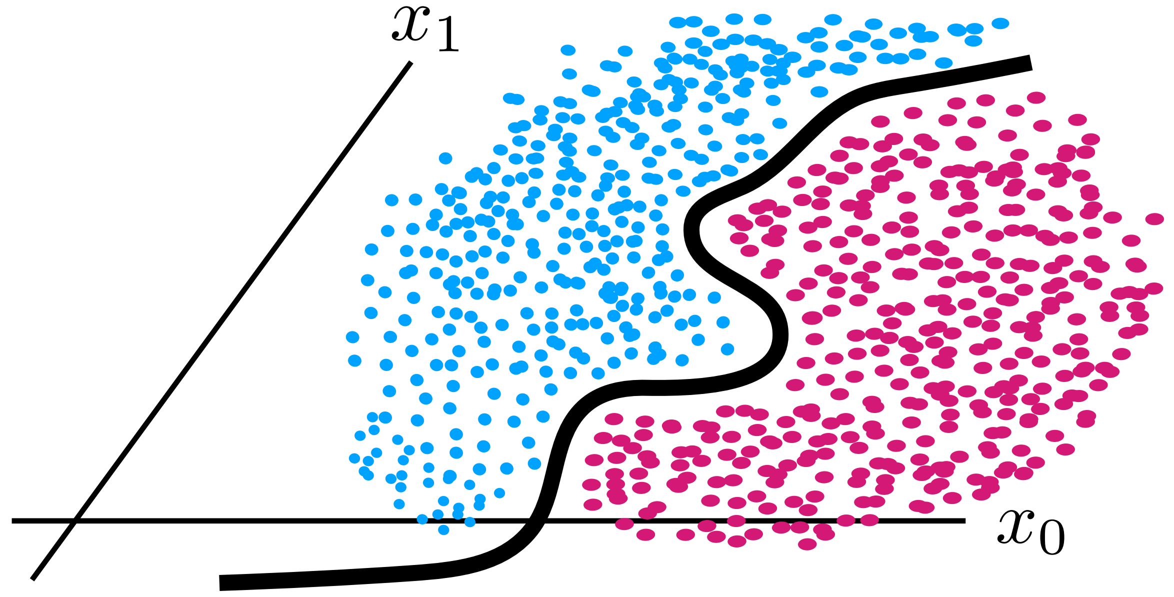
Support Vector Machine (SVM)

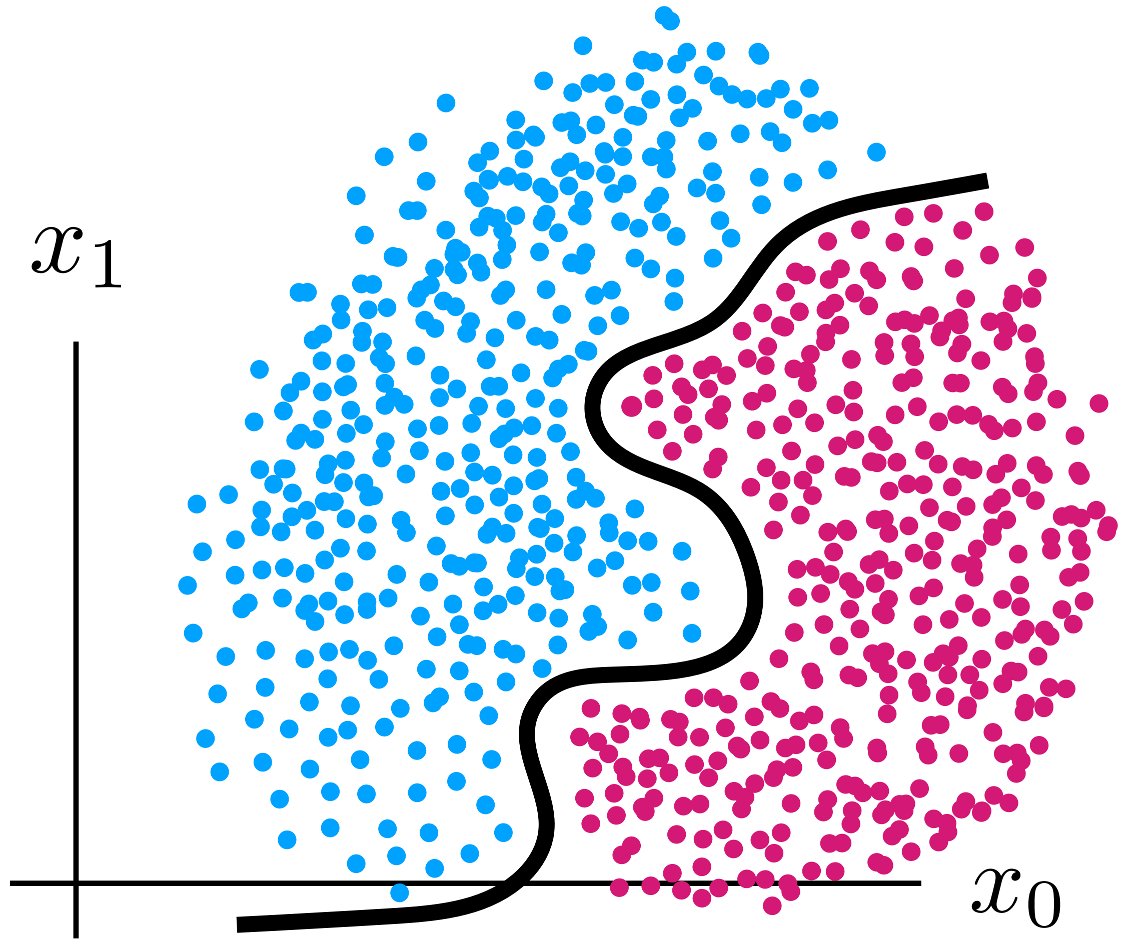
Logistic Regression (smoothed)

COST: $$\min_{\theta} \quad \lambda||\theta_{1:n}||_2^2 + \ \frac{1}{T}\sum_t \max\left\{0, 1 - \gamma_t(\theta_{1:n}^T h_t(x_t) - \theta_0)\right\}$$

Soft boundary







Z

OUTPUTS (Dependent Variables)

$$\gamma_t = \mathrm{sgn}(\theta^T x_t - \theta_0)$$

INPUTS (Independent Variables)

$$\begin{bmatrix} \boxed{\begin{matrix}\gamma_{00}\\ \gamma_{10}\\ \gamma_{20}\\ \gamma_{30}\\ \gamma_{40}\\ \vdots\\ \gamma_{T0}\end{matrix}} & \begin{matrix}\cdots\\ \cdots\\ \cdots\\ \cdots\\ \cdots\\ \\ \cdots\end{matrix} & \begin{matrix}\gamma_{0m'}\\ \gamma_{1m'}\\ \gamma_{2m'}\\ \gamma_{3m'}\\ \gamma_{4m'}\\ \vdots\\ \gamma_{Tm'}\end{matrix} & \begin{matrix}y_{00}\\ y_{10}\\ y_{20}\\ y_{30}\\ y_{40}\\ \vdots\\ y_{T0}\end{matrix} & \begin{matrix}\cdots\\ \cdots\\ \cdots\\ \cdots\\ \cdots\\ \\ \cdots\end{matrix} & \begin{matrix}y_{0m}\\ y_{1m}\\ y_{2m}\\ y_{3m}\\ y_{4m}\\ \vdots\\ y_{Tm}\end{matrix} \end{bmatrix} \Leftarrow \begin{bmatrix} h_{00}(x_0) & \cdots & h_{0n}(x_0)\\ h_{10}(x_1) & \cdots & h_{1n}(x_1)\\ h_{20}(x_2) & \cdots & h_{2n}(x_2)\\ h_{30}(x_3) & \cdots & h_{3n}(x_3)\\ h_{40}(x_4) & \cdots & h_{4n}(x_4)\\ \vdots & & \vdots\\ h_{T0}(x_T) & \cdots & h_{Tn}(x_T) \end{bmatrix} \Leftarrow \begin{bmatrix} x_{00} & \cdots & x_{0n}\\ x_{10} & \cdots & x_{1n}\\ x_{20} & \cdots & x_{2n}\\ x_{30} & \cdots & x_{3n}\\ x_{40} & \cdots & x_{4n}\\ \vdots & & \vdots\\ x_{T0} & \cdots & x_{Tn} \end{bmatrix}$$

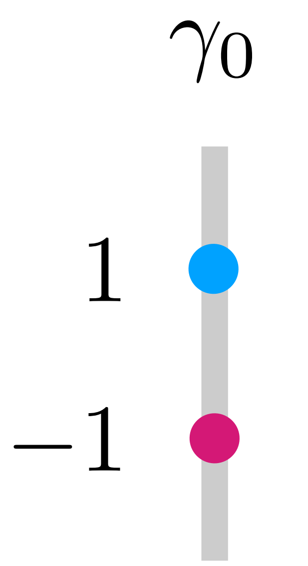
"Binary classifier"

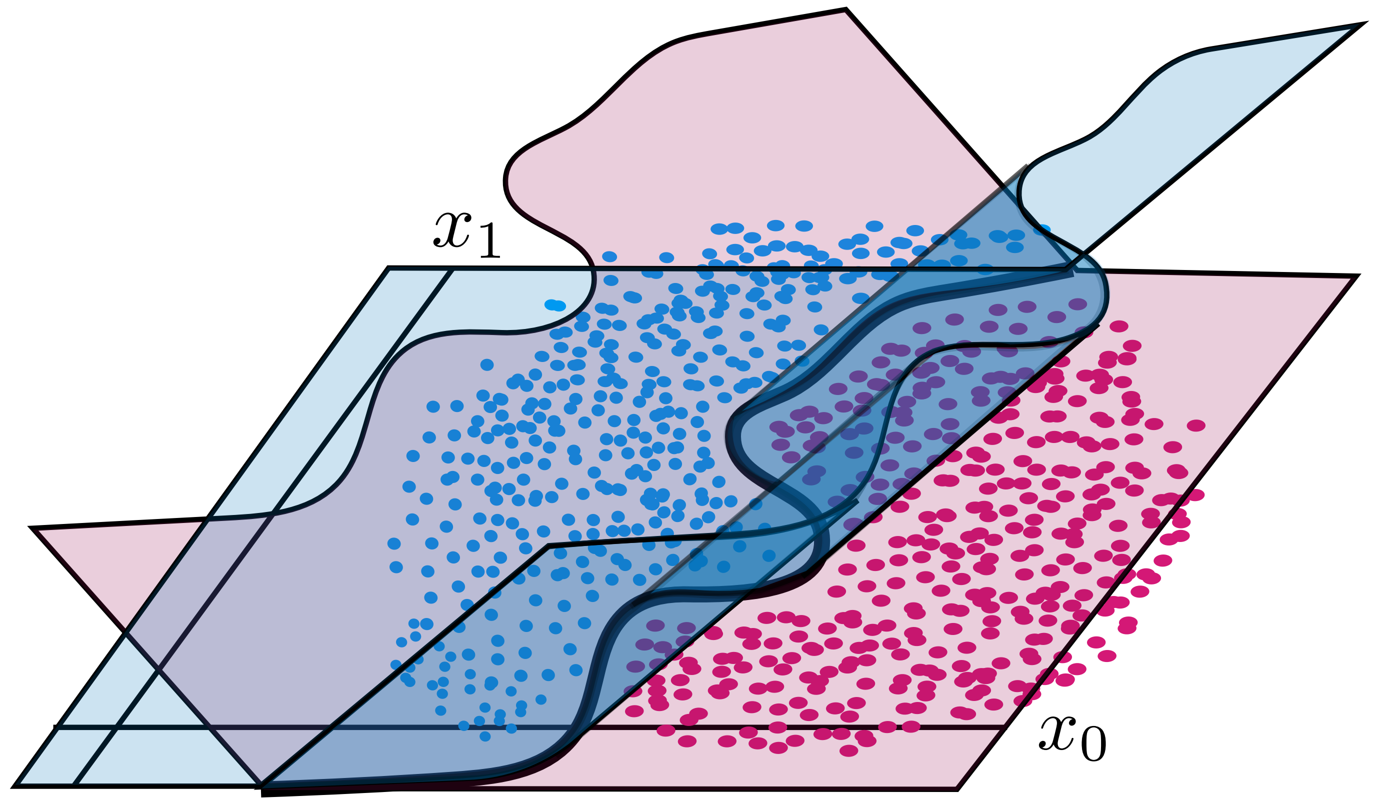
Support Vector Machine (SVM)

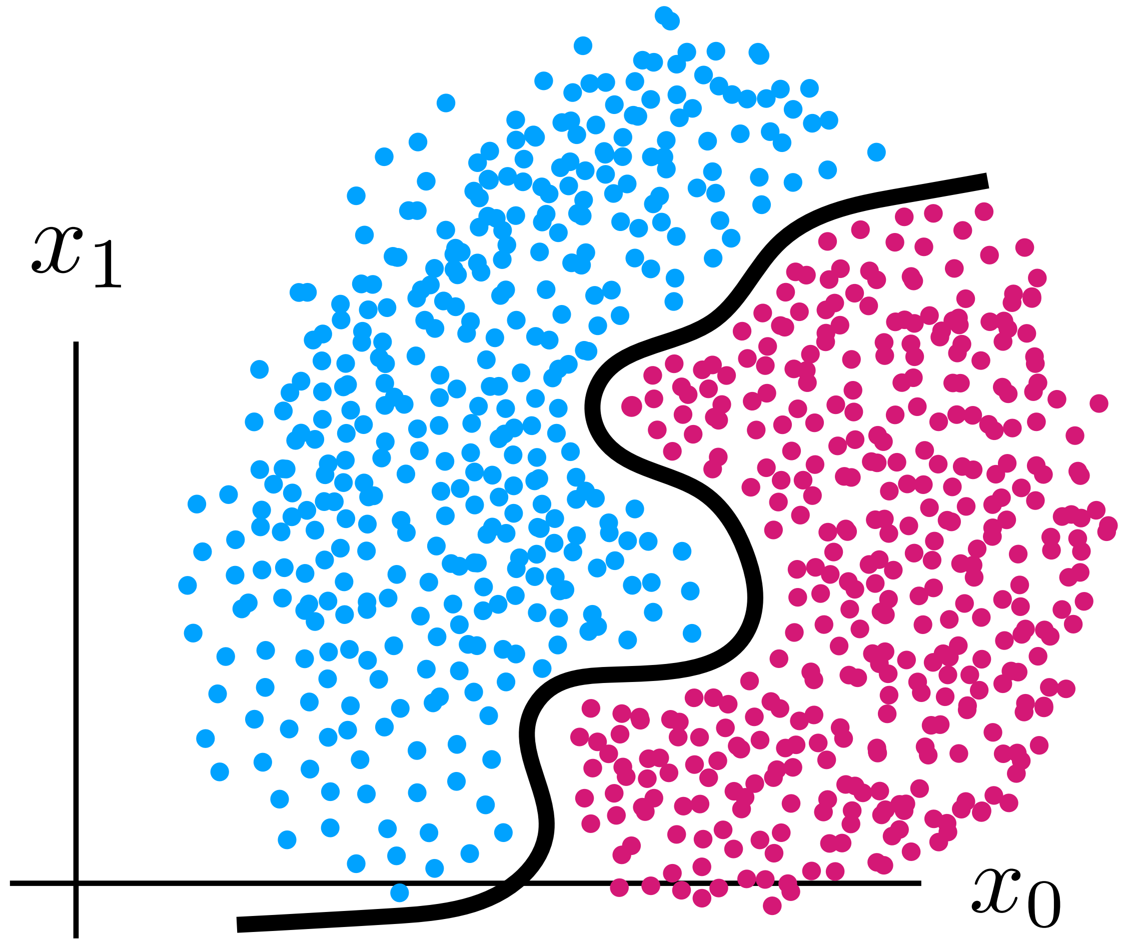
Logistic Regression (smoothed)

COST: $$\min_{\theta} \lambda||\theta_{1:n}||_2^2 + \frac{1}{T}\sum_t \max\{0, 1-\gamma_t(\theta_{1:n}^T h_t(x_t) - \theta_0)\}$$

Soft boundary







Classification

OUTPUTS

(Dependent Variables)

$$\gamma_t = \mathrm{sgn}(\theta^T x_t - \theta_0)$$

INPUTS

(Independent Variables)

$$\begin{bmatrix} \boxed{\begin{matrix} \gamma_{00} \\ \gamma_{10} \\ \gamma_{20} \\ \gamma_{30} \\ \gamma_{40} \\ \vdots \\ \gamma_{T0} \end{matrix}} & \begin{matrix} \cdots \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ \\ \cdots \end{matrix} & \begin{matrix} \gamma_{0m'} \\ \gamma_{1m'} \\ \gamma_{2m'} \\ \gamma_{3m'} \\ \gamma_{4m'} \\ \vdots \\ \gamma_{Tm'} \end{matrix} & \begin{matrix} y_{00} \\ y_{10} \\ y_{20} \\ y_{30} \\ y_{40} \\ \vdots \\ y_{T0} \end{matrix} & \begin{matrix} \cdots \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ \\ \cdots \end{matrix} & \begin{matrix} y_{0m} \\ y_{1m} \\ y_{2m} \\ y_{3m} \\ y_{4m} \\ \vdots \\ y_{Tm} \end{matrix} \end{bmatrix} \Leftarrow \begin{bmatrix} h_{00}(x_0) & \cdots & h_{0n}(x_0) \\ h_{10}(x_1) & \cdots & h_{1n}(x_1) \\ h_{20}(x_2) & \cdots & h_{2n}(x_2) \\ h_{30}(x_3) & \cdots & h_{3n}(x_3) \\ h_{40}(x_4) & \cdots & h_{4n}(x_4) \\ \vdots & & \vdots \\ h_{T0}(x_T) & \cdots & h_{Tn}(x_T) \end{bmatrix} \Leftarrow \begin{bmatrix} x_{00} & \cdots & x_{0n} \\ x_{10} & \cdots & x_{1n} \\ x_{20} & \cdots & x_{2n} \\ x_{30} & \cdots & x_{3n} \\ x_{40} & \cdots & x_{4n} \\ \vdots & & \vdots \\ x_{T0} & \cdots & x_{Tn} \end{bmatrix}$$

Collection of Binary Classifiers

"One to the others"

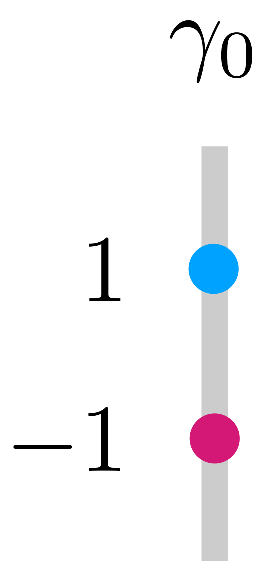
# of classifiers: m

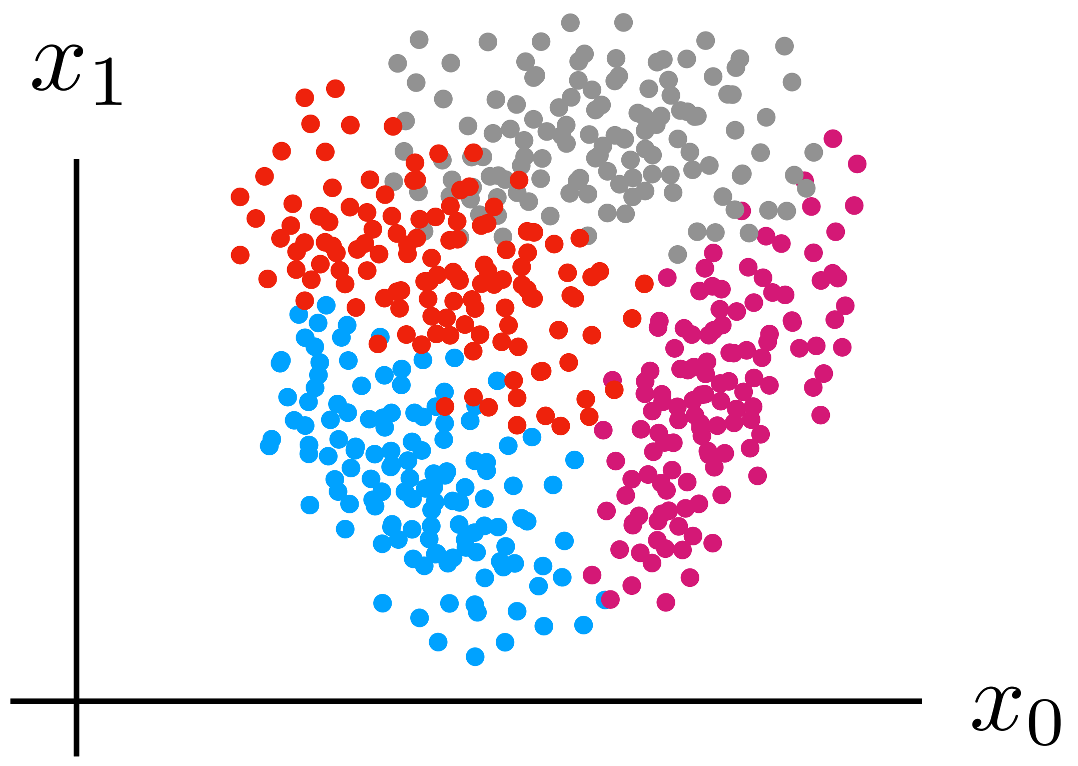
Support Vector Machine (SVM)

Logistic Regression (smoothed)

COST: $$\min_{\theta} \quad \lambda||\theta_{1:n}||_2^2 + \frac{1}{T}\sum_t \max\{0, 1 - \gamma_t(\theta_{1:n}^T h_t(x_t) - \theta_0)\}$$

Soft boundary





Classification

OUTPUTS

(Dependent Variables)

$$\gamma_t = \mathrm{sgn}(\theta^T x_t - \theta_0)$$

INPUTS

(Independent Variables)

$$\begin{bmatrix} \gamma_{00} & \cdots & \gamma_{0m'} & y_{00} & \cdots & y_{0m} \\ \gamma_{10} & \cdots & \gamma_{1m'} & y_{10} & \cdots & y_{1m} \\ \gamma_{20} & \cdots & \gamma_{2m'} & y_{20} & \cdots & y_{2m} \\ \gamma_{30} & \cdots & \gamma_{3m'} & y_{30} & \cdots & y_{3m} \\ \gamma_{40} & \cdots & \gamma_{4m'} & y_{40} & \cdots & y_{4m} \\ \vdots & & \vdots & \vdots & & \vdots \\ \gamma_{T0} & \cdots & \gamma_{Tm'} & y_{T0} & \cdots & y_{Tm} \end{bmatrix} \Leftarrow \begin{bmatrix} h_{00}(x_0) & \cdots & h_{0n}(x_0) \\ h_{10}(x_1) & \cdots & h_{1n}(x_1) \\ h_{20}(x_2) & \cdots & h_{2n}(x_2) \\ h_{30}(x_3) & \cdots & h_{3n}(x_3) \\ h_{40}(x_4) & \cdots & h_{4n}(x_4) \\ \vdots & & \vdots \\ h_{T0}(x_T) & \cdots & h_{Tn}(x_T) \end{bmatrix} \Leftarrow \begin{bmatrix} x_{00} & \cdots & x_{0n} \\ x_{10} & \cdots & x_{1n} \\ x_{20} & \cdots & x_{2n} \\ x_{30} & \cdots & x_{3n} \\ x_{40} & \cdots & x_{4n} \\ \vdots & & \vdots \\ x_{T0} & \cdots & x_{Tn} \end{bmatrix}$$

Collection of Binary Classifiers

"One to the others"

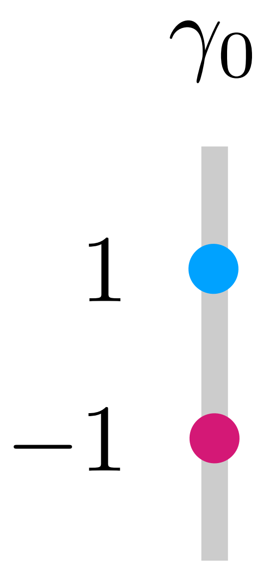
# of classifiers: m

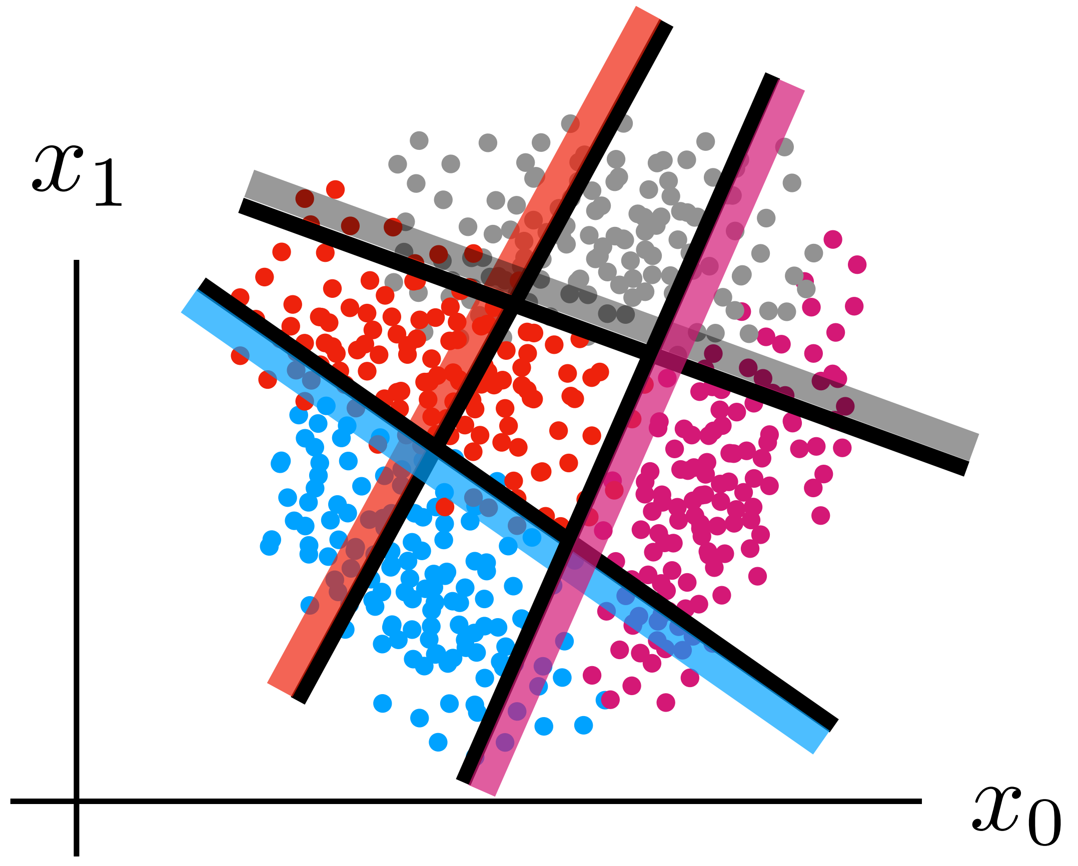
Support Vector Machine (SVM)

Logistic Regression (smoothed)

COST: $$\min_{\theta} \lambda||\theta_{1:n}||_2^2 + \frac{1}{T}\sum_t \max\{0, 1 - \gamma_t(\theta_{1:n}^T h_t(x_t) - \theta_0)\}$$

Soft boundary





Classification

OUTPUTS

(Dependent Variables)

$$\gamma_t = \mathrm{sgn}(\theta^T x_t - \theta_0)$$

INPUTS

(Independent Variables)

$$\begin{bmatrix} \gamma_{00} & \cdots & \gamma_{0m'} & y_{00} & \cdots & y_{0m} \\ \gamma_{10} & \cdots & \gamma_{1m'} & y_{10} & \cdots & y_{1m} \\ \gamma_{20} & \cdots & \gamma_{2m'} & y_{20} & \cdots & y_{2m} \\ \gamma_{30} & \cdots & \gamma_{3m'} & y_{30} & \cdots & y_{3m} \\ \gamma_{40} & \cdots & \gamma_{4m'} & y_{40} & \cdots & y_{4m} \\ \vdots & & \vdots & \vdots & & \vdots \\ \gamma_{T0} & \cdots & \gamma_{Tm'} & y_{T0} & \cdots & y_{Tm} \end{bmatrix} \Leftarrow \begin{bmatrix} h_{00}(x_0) & \cdots & h_{0n}(x_0) \\ h_{10}(x_1) & \cdots & h_{1n}(x_1) \\ h_{20}(x_2) & \cdots & h_{2n}(x_2) \\ h_{30}(x_3) & \cdots & h_{3n}(x_3) \\ h_{40}(x_4) & \cdots & h_{4n}(x_4) \\ \vdots & & \vdots \\ h_{T0}(x_T) & \cdots & h_{Tn}(x_T) \end{bmatrix} \Leftarrow \begin{bmatrix} x_{00} & \cdots & x_{0n} \\ x_{10} & \cdots & x_{1n} \\ x_{20} & \cdots & x_{2n} \\ x_{30} & \cdots & x_{3n} \\ x_{40} & \cdots & x_{4n} \\ \vdots & & \vdots \\ x_{T0} & \cdots & x_{Tn} \end{bmatrix}$$

Collection of Binary Classifiers

"One to the others"

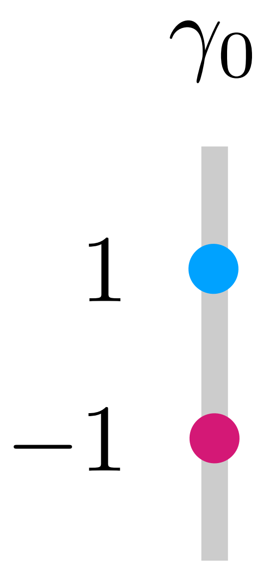
# of classifiers: m

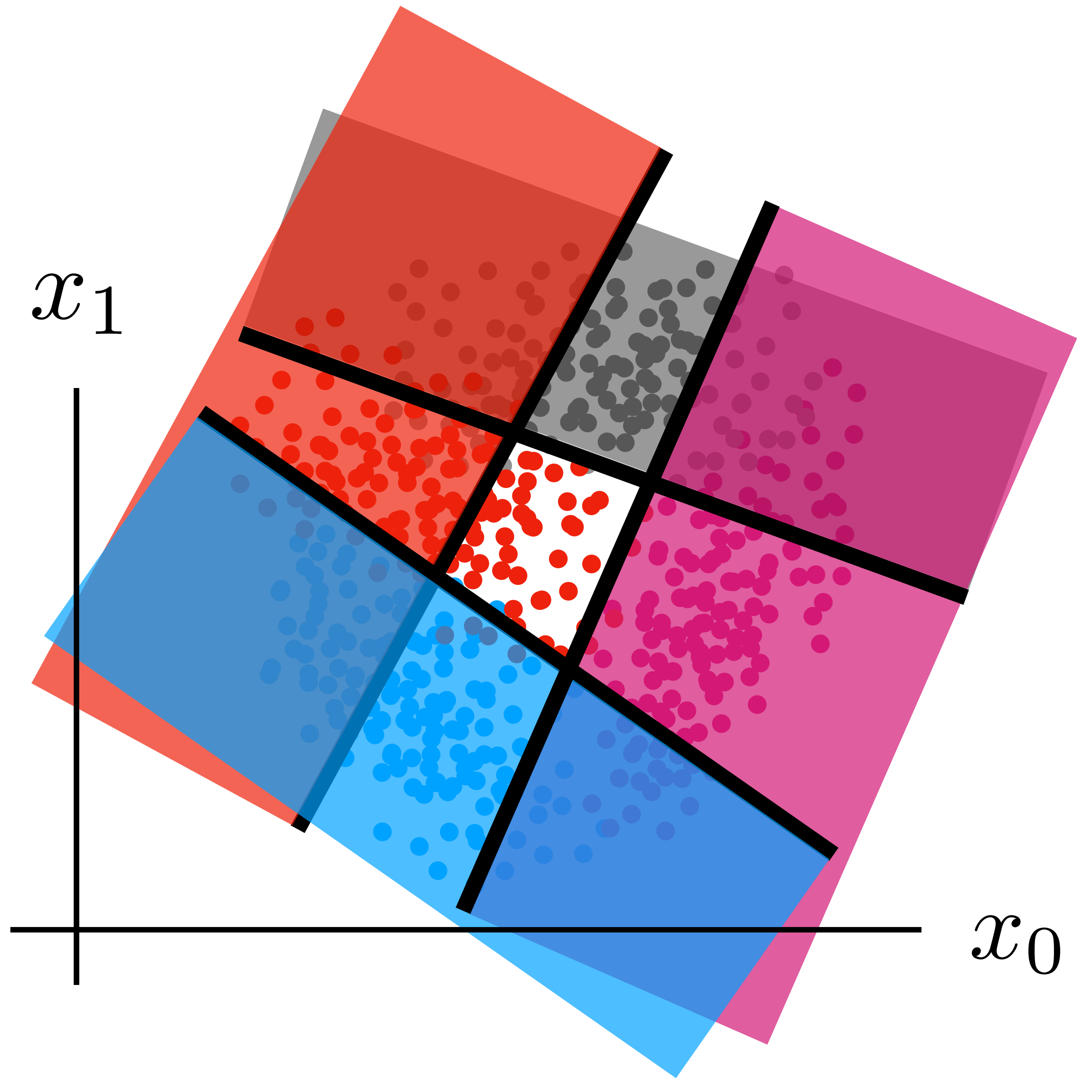
Support Vector Machine (SVM)

Logistic Regression (smoothed)

COST: $\min_{\theta} \; \lambda||\theta_{1:n}||_2^2 + \frac{1}{T}\sum_t \max\left\{0, 1 - \gamma_t(\theta_{1:n}^T h_t(x_t) - \theta_0)\right\}$

Soft boundary





Classification

OUTPUTS
(Dependent Variables)

$$\gamma_t = \mathrm{sgn}(\theta^T x_t - \theta_0)$$

INPUTS
(Independent Variables)

$$\left[\begin{array}{ccccccc} \boxed{\begin{array}{c}\gamma_{00} \\ \gamma_{10} \\ \gamma_{20} \\ \gamma_{30} \\ \gamma_{40} \\ \vdots \\ \gamma_{T0}\end{array}} & \begin{array}{c}\cdots \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ \\ \cdots\end{array} & \begin{array}{c}\gamma_{0m'} \\ \gamma_{1m'} \\ \gamma_{2m'} \\ \gamma_{3m'} \\ \gamma_{4m'} \\ \vdots \\ \gamma_{Tm'}\end{array} & \begin{array}{c}y_{00} \\ y_{10} \\ y_{20} \\ y_{30} \\ y_{40} \\ \vdots \\ y_{T0}\end{array} & \begin{array}{c}\cdots \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ \\ \cdots\end{array} & \begin{array}{c}y_{0m} \\ y_{1m} \\ y_{2m} \\ y_{3m} \\ y_{4m} \\ \vdots \\ y_{Tm}\end{array} \end{array}\right] \Leftarrow \begin{bmatrix} h_{00}(x_0) & \cdots & h_{0n}(x_0) \\ h_{10}(x_1) & \cdots & h_{1n}(x_1) \\ h_{20}(x_2) & \cdots & h_{2n}(x_2) \\ h_{30}(x_3) & \cdots & h_{3n}(x_3) \\ h_{40}(x_4) & \cdots & h_{4n}(x_4) \\ \vdots & & \vdots \\ h_{T0}(x_T) & \cdots & h_{Tn}(x_T) \end{bmatrix} \Leftarrow \begin{bmatrix} x_{00} & \cdots & x_{0n} \\ x_{10} & \cdots & x_{1n} \\ x_{20} & \cdots & x_{2n} \\ x_{30} & \cdots & x_{3n} \\ x_{40} & \cdots & x_{4n} \\ \vdots & & \vdots \\ x_{T0} & \cdots & x_{Tn} \end{bmatrix}$$

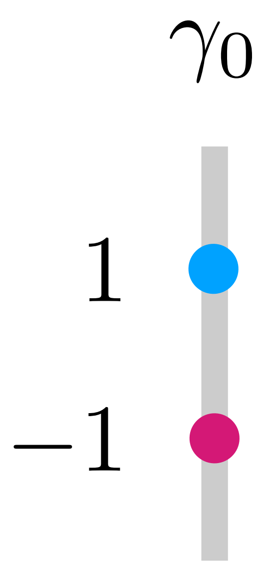
Collection of Binary Classifiers
Pairwise - "One to another"
# of classifiers: m(m+1)/2

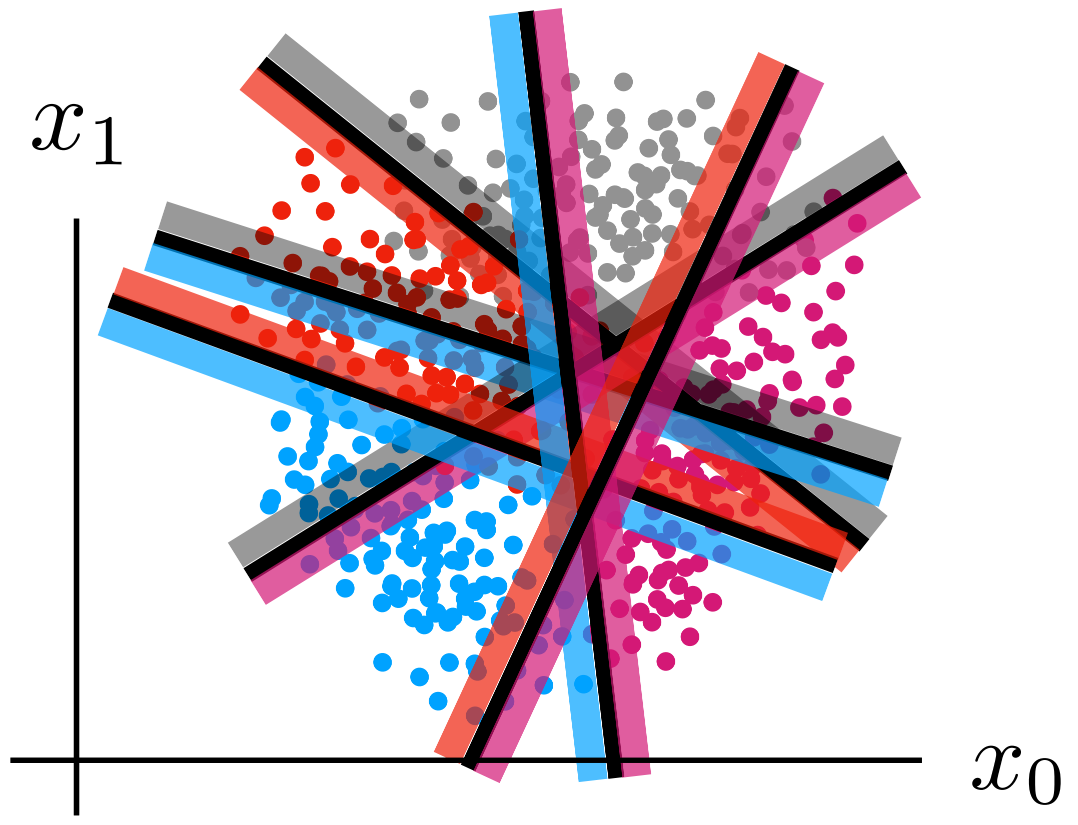
Support Vector Machine (SVM)

Logistic Regression (smoothed)

COST: $$\min_{\theta} \lambda||\theta_{1:n}||_2^2 + \frac{1}{T}\sum_t \max\left\{0, 1 - \gamma_t(\theta_{1:n}^T h_t(x_t) - \theta_0)\right\}$$

Soft boundary





Classification

OUTPUTS

(Dependent Variables)

$$\gamma_t = \mathrm{sgn}(\theta^T x_t - \theta_0)$$

INPUTS

(Independent Variables)

$$\begin{bmatrix} \boxed{\begin{matrix}\gamma_{00} \\ \gamma_{10} \\ \gamma_{20} \\ \gamma_{30} \\ \gamma_{40} \\ \vdots \\ \gamma_{T0}\end{matrix}} & \begin{matrix}\cdots \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ \\ \cdots\end{matrix} & \begin{matrix}\gamma_{0m'} \\ \gamma_{1m'} \\ \gamma_{2m'} \\ \gamma_{3m'} \\ \gamma_{4m'} \\ \vdots \\ \gamma_{Tm'}\end{matrix} & \begin{matrix}y_{00} \\ y_{10} \\ y_{20} \\ y_{30} \\ y_{40} \\ \vdots \\ y_{T0}\end{matrix} & \begin{matrix}\cdots \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ \\ \cdots\end{matrix} & \begin{matrix}y_{0m} \\ y_{1m} \\ y_{2m} \\ y_{3m} \\ y_{4m} \\ \vdots \\ y_{Tm}\end{matrix} \end{bmatrix} \Leftarrow \begin{bmatrix} h_{00}(x_0) & \cdots & h_{0n}(x_0) \\ h_{10}(x_1) & \cdots & h_{1n}(x_1) \\ h_{20}(x_2) & \cdots & h_{2n}(x_2) \\ h_{30}(x_3) & \cdots & h_{3n}(x_3) \\ h_{40}(x_4) & \cdots & h_{4n}(x_4) \\ \vdots & & \vdots \\ h_{T0}(x_T) & \cdots & h_{Tn}(x_T) \end{bmatrix} \Leftarrow \begin{bmatrix} x_{00} & \cdots & x_{0n} \\ x_{10} & \cdots & x_{1n} \\ x_{20} & \cdots & x_{2n} \\ x_{30} & \cdots & x_{3n} \\ x_{40} & \cdots & x_{4n} \\ \vdots & & \vdots \\ x_{T0} & \cdots & x_{Tn} \end{bmatrix}$$

Collection of Binary Classifiers

Pairwise - “One to another”

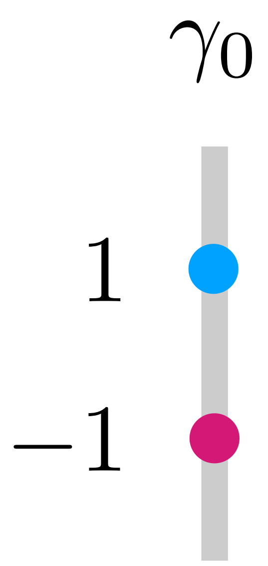
# of classifiers: m(m+1)/2

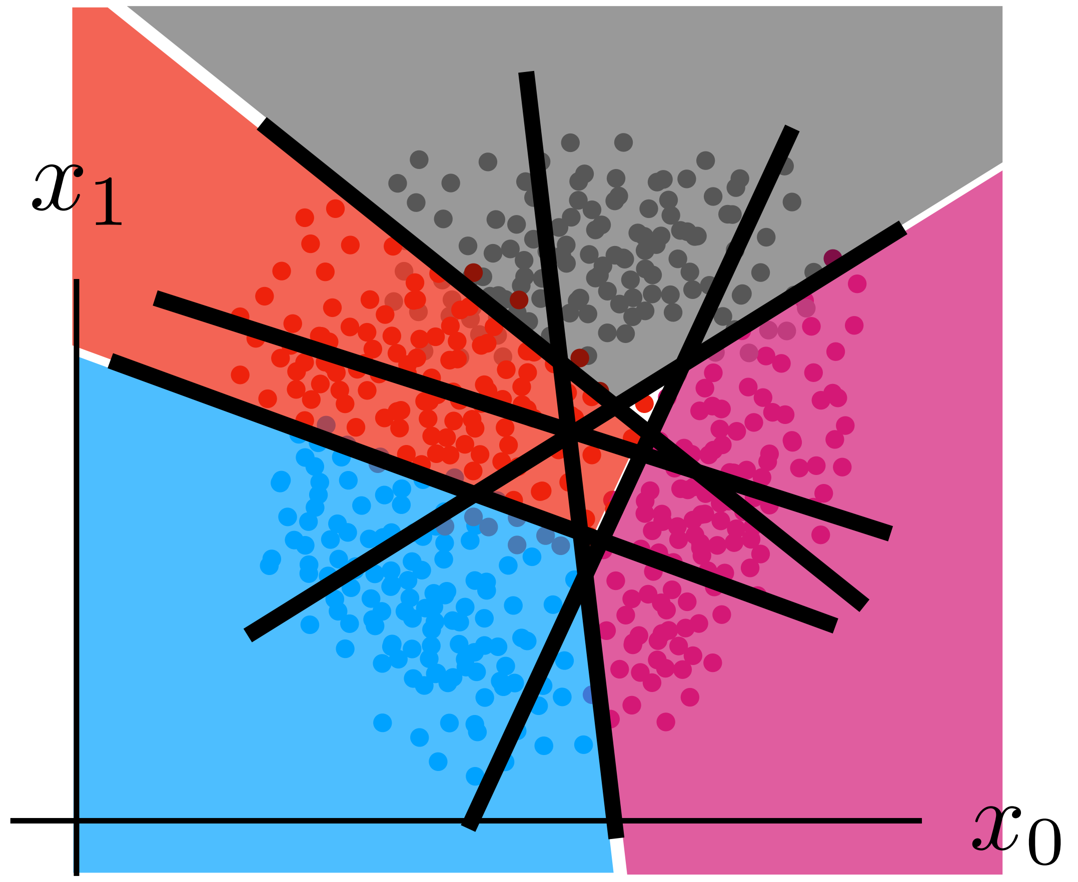
Support Vector Machine (SVM)

Logistic Regression (smoothed)

COST: $$\min_{\theta} \lambda||\theta_{1:n}||_2^2 + \frac{1}{T}\sum_t \max\{0, 1-\gamma_t(\theta_{1:n}^T h_t(x_t)-\theta_0)\}$$

Soft boundary





Classification

OUTPUTS

(Dependent Variables)

$$\gamma_t = \mathrm{sgn}(\theta^T x_t - \theta_0)$$

INPUTS

(Independent Variables)

$$\left[\begin{array}{ccc|ccc} \boxed{\begin{matrix}\gamma_{00}\\ \gamma_{10}\\ \gamma_{20}\\ \gamma_{30}\\ \gamma_{40}\\ \vdots\\ \gamma_{T0}\end{matrix}} & \begin{matrix}\cdots\\ \cdots\\ \cdots\\ \cdots\\ \cdots\\ \\ \cdots\end{matrix} & \begin{matrix}\gamma_{0m'}\\ \gamma_{1m'}\\ \gamma_{2m'}\\ \gamma_{3m'}\\ \gamma_{4m'}\\ \vdots\\ \gamma_{Tm'}\end{matrix} & \begin{matrix}y_{00}\\ y_{10}\\ y_{20}\\ y_{30}\\ y_{40}\\ \vdots\\ y_{T0}\end{matrix} & \begin{matrix}\cdots\\ \cdots\\ \cdots\\ \cdots\\ \cdots\\ \\ \cdots\end{matrix} & \begin{matrix}y_{0m}\\ y_{1m}\\ y_{2m}\\ y_{3m}\\ y_{4m}\\ \vdots\\ y_{Tm}\end{matrix} \end{array}\right] \Leftarrow \begin{bmatrix} h_{00}(x_0) & \cdots & h_{0n}(x_0)\\ h_{10}(x_1) & \cdots & h_{1n}(x_1)\\ h_{20}(x_2) & \cdots & h_{2n}(x_2)\\ h_{30}(x_3) & \cdots & h_{3n}(x_3)\\ h_{40}(x_4) & \cdots & h_{4n}(x_4)\\ \vdots & & \vdots\\ h_{T0}(x_T) & \cdots & h_{Tn}(x_T) \end{bmatrix} \Leftarrow \begin{bmatrix} x_{00} & \cdots & x_{0n}\\ x_{10} & \cdots & x_{1n}\\ x_{20} & \cdots & x_{2n}\\ x_{30} & \cdots & x_{3n}\\ x_{40} & \cdots & x_{4n}\\ \vdots & & \vdots\\ x_{T0} & \cdots & x_{Tn} \end{bmatrix}$$

Collection of Binary Classifiers

Pairwise - "One to another"

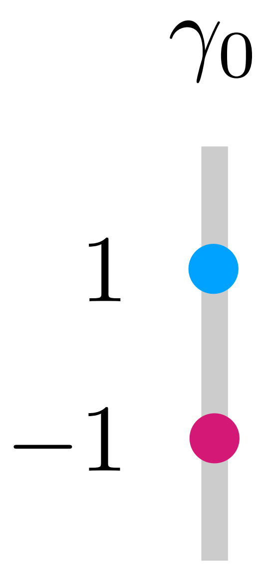
# of classifiers: m(m+1)/2

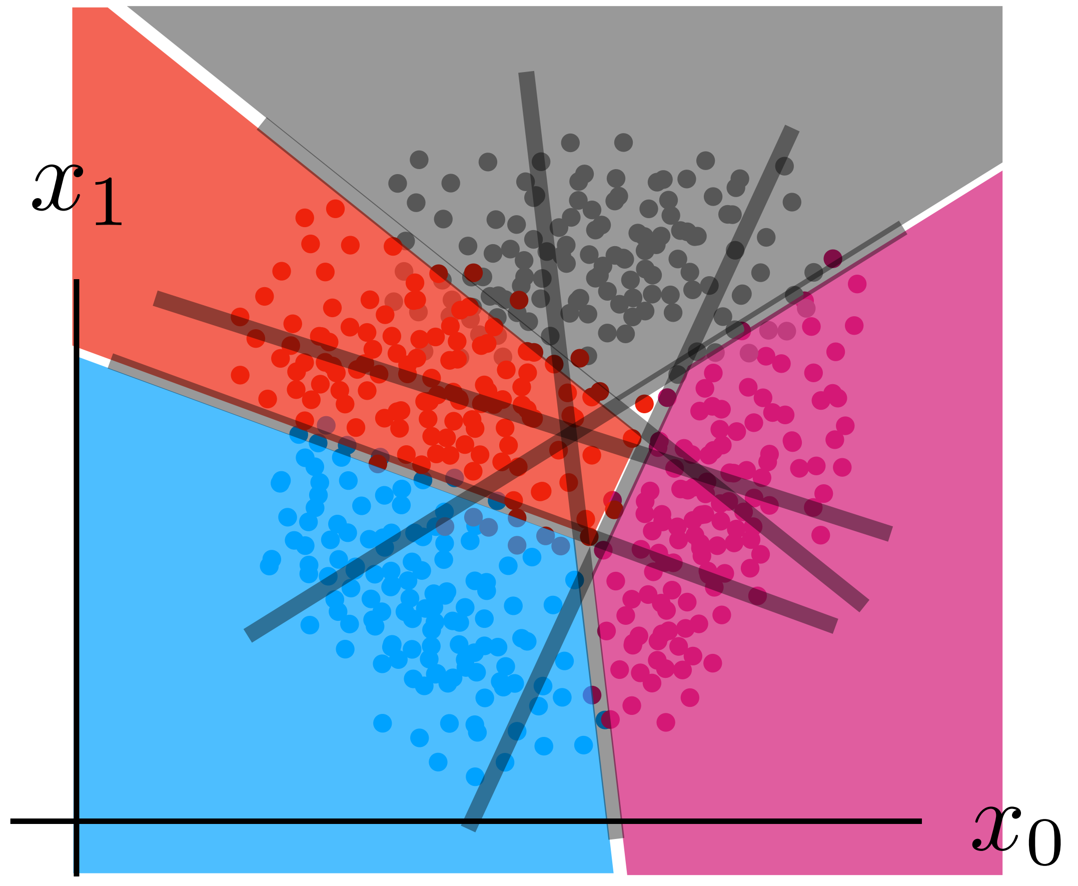
Support Vector Machine (SVM)

Logistic Regression (smoothed)

COST: $$\min_{\theta} \; \lambda||\theta_{1:n}||_2^2 + \frac{1}{T}\sum_t \max\left\{0, 1 - \gamma_t(\theta_{1:n}^T h_t(x_t) - \theta_0)\right\}$$

Soft boundary





Classification

OUTPUTS
(Dependent Variables)

$$\gamma_t = \mathrm{sgn}(\theta^T x_t - \theta_0)$$

INPUTS
(Independent Variables)

$$\begin{bmatrix} \gamma_{00} & \cdots & \gamma_{0m'} & y_{00} & \cdots & y_{0m} \\ \gamma_{10} & \cdots & \gamma_{1m'} & y_{10} & \cdots & y_{1m} \\ \gamma_{20} & \cdots & \gamma_{2m'} & y_{20} & \cdots & y_{2m} \\ \gamma_{30} & \cdots & \gamma_{3m'} & y_{30} & \cdots & y_{3m} \\ \gamma_{40} & \cdots & \gamma_{4m'} & y_{40} & \cdots & y_{4m} \\ \vdots & & \vdots & \vdots & & \vdots \\ \gamma_{T0} & \cdots & \gamma_{Tm'} & y_{T0} & \cdots & y_{Tm} \end{bmatrix} \Leftarrow \begin{bmatrix} h_{00}(x_0) & \cdots & h_{0n}(x_0) \\ h_{10}(x_1) & \cdots & h_{1n}(x_1) \\ h_{20}(x_2) & \cdots & h_{2n}(x_2) \\ h_{30}(x_3) & \cdots & h_{3n}(x_3) \\ h_{40}(x_4) & \cdots & h_{4n}(x_4) \\ \vdots & & \vdots \\ h_{T0}(x_T) & \cdots & h_{Tn}(x_T) \end{bmatrix} \Leftarrow \begin{bmatrix} x_{00} & \cdots & x_{0n} \\ x_{10} & \cdots & x_{1n} \\ x_{20} & \cdots & x_{2n} \\ x_{30} & \cdots & x_{3n} \\ x_{40} & \cdots & x_{4n} \\ \vdots & & \vdots \\ x_{T0} & \cdots & x_{Tn} \end{bmatrix}$$

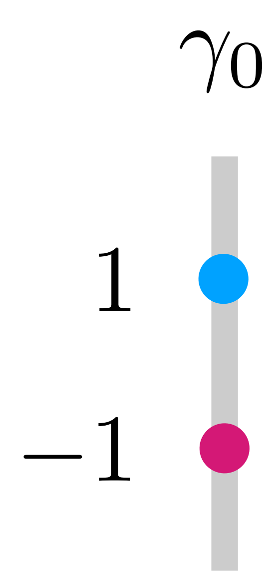
Collection of Binary Classifiers
Pairwise - "One to another"
# of classifiers: m(m+1)/2

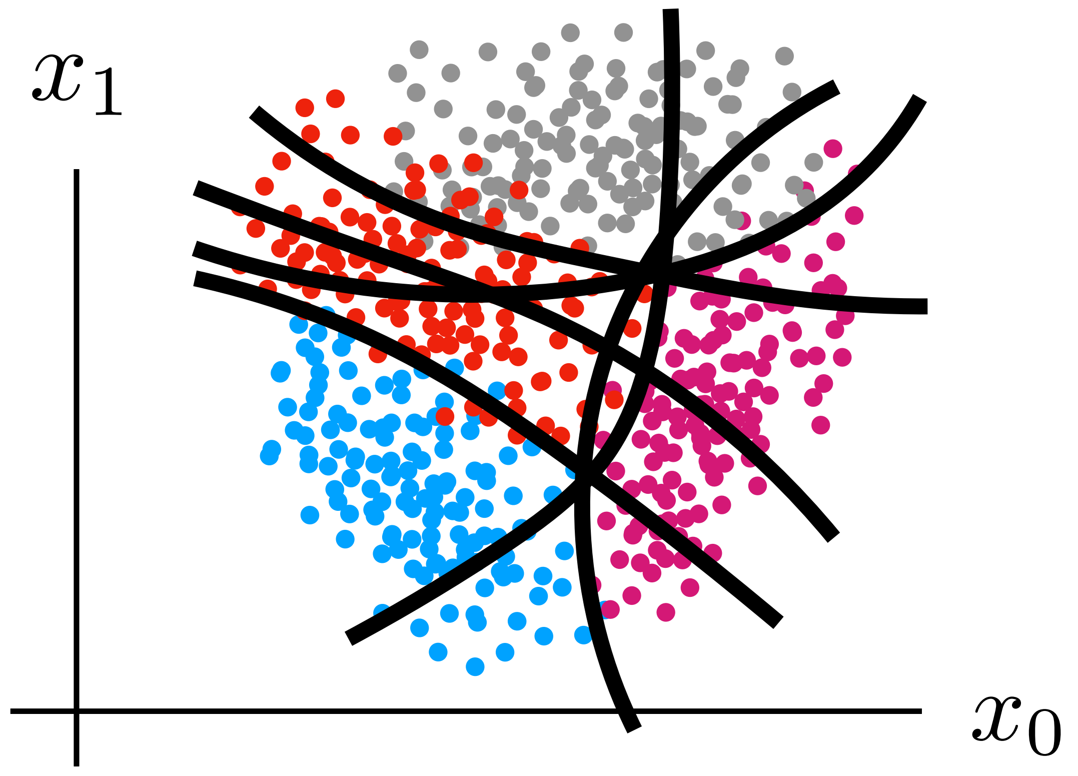
Support Vector Machine (SVM)

Logistic Regression (smoothed)

COST: $$\min_{\theta} \lambda||\theta_{1:n}||_2^2 + \frac{1}{T}\sum_t \max\{0, 1 - \gamma_t(\theta_{1:n}^T h_t(x_t) - \theta_0)\}$$

Soft boundary





Classification

OUTPUTS

(Dependent Variables)

$$\gamma_t = \mathrm{sgn}(\theta^T x_t - \theta_0)$$

INPUTS

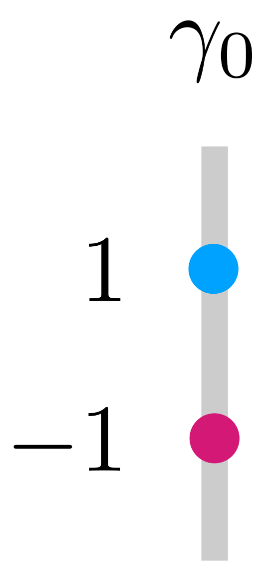
(Independent Variables)

$$\begin{bmatrix} \boxed{\begin{matrix} \gamma_{00} \\ \gamma_{10} \\ \gamma_{20} \\ \gamma_{30} \\ \gamma_{40} \\ \vdots \\ \gamma_{T0} \end{matrix}} & \begin{matrix} \cdots \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ \\ \cdots \end{matrix} & \begin{matrix} \gamma_{0m'} \\ \gamma_{1m'} \\ \gamma_{2m'} \\ \gamma_{3m'} \\ \gamma_{4m'} \\ \vdots \\ \gamma_{Tm'} \end{matrix} & \begin{matrix} y_{00} \\ y_{10} \\ y_{20} \\ y_{30} \\ y_{40} \\ \vdots \\ y_{T0} \end{matrix} & \begin{matrix} \cdots \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ \\ \cdots \end{matrix} & \begin{matrix} y_{0m} \\ y_{1m} \\ y_{2m} \\ y_{3m} \\ y_{4m} \\ \vdots \\ y_{Tm} \end{matrix} \end{bmatrix} \Leftarrow \begin{bmatrix} h_{00}(x_0) & \cdots & h_{0n}(x_0) \\ h_{10}(x_1) & \cdots & h_{1n}(x_1) \\ h_{20}(x_2) & \cdots & h_{2n}(x_2) \\ h_{30}(x_3) & \cdots & h_{3n}(x_3) \\ h_{40}(x_4) & \cdots & h_{4n}(x_4) \\ \vdots & & \vdots \\ h_{T0}(x_T) & \cdots & h_{Tn}(x_T) \end{bmatrix} \Leftarrow \begin{bmatrix} x_{00} & \cdots & x_{0n} \\ x_{10} & \cdots & x_{1n} \\ x_{20} & \cdots & x_{2n} \\ x_{30} & \cdots & x_{3n} \\ x_{40} & \cdots & x_{4n} \\ \vdots & & \vdots \\ x_{T0} & \cdots & x_{Tn} \end{bmatrix}$$

Collection of Binary Classifiers

Pairwise - "One to another"

# of classifiers: m(m+1)/2

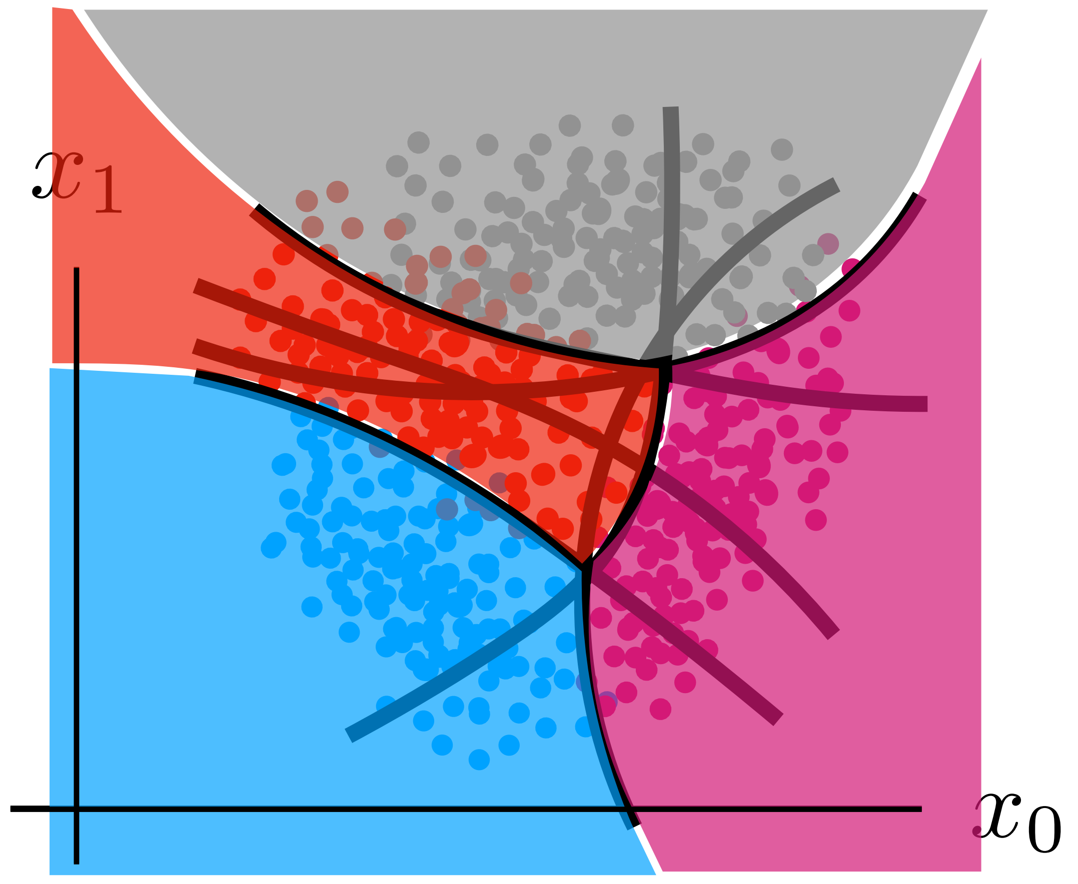


Support Vector Machine (SVM)

Logistic Regression (smoothed)

COST: $\min_{\theta} \; \lambda||\theta_{1:n}||_2^2 + \frac{1}{T}\sum_t \max\{0, 1 - \gamma_t(\theta_{1:n}^T h_t(x_t) - \theta_0)\}$

Soft boundary



Discriminant Analysis

OUTPUTS (Dependent Variables) $\Leftarrow$ **INPUTS (Independent Variables)**

$$\begin{bmatrix} \boxed{\begin{matrix}\gamma_{00}\\ \gamma_{10}\\ \gamma_{20}\\ \gamma_{30}\\ \gamma_{40}\\ \vdots \\ \gamma_{T0}\end{matrix}} & \begin{matrix}\cdots\\ \cdots\\ \cdots\\ \cdots\\ \cdots\\ \\ \cdots\end{matrix} & \begin{matrix}\gamma_{0m'}\\ \gamma_{1m'}\\ \gamma_{2m'}\\ \gamma_{3m'}\\ \gamma_{4m'}\\ \vdots \\ \gamma_{Tm'}\end{matrix} & \begin{matrix}y_{00}\\ y_{10}\\ y_{20}\\ y_{30}\\ y_{40}\\ \vdots \\ y_{T0}\end{matrix} & \begin{matrix}\cdots\\ \cdots\\ \cdots\\ \cdots\\ \cdots\\ \\ \cdots\end{matrix} & \begin{matrix}y_{0m}\\ y_{1m}\\ y_{2m}\\ y_{3m}\\ y_{4m}\\ \vdots \\ y_{Tm}\end{matrix} \end{bmatrix} \Leftarrow \boxed{f\left[\begin{pmatrix} x_{00} & \cdots & x_{0n} \\ x_{10} & \cdots & x_{1n} \end{pmatrix}\right.} \begin{matrix} x_{20} & \cdots & x_{2n} \\ x_{30} & \cdots & x_{3n} \\ x_{40} & \cdots & x_{4n} \\ \vdots & & \vdots \\ x_{T0} & \cdots & x_{Tn} \end{matrix} \Bigg] \begin{bmatrix} \xi_{00} & \cdots & \xi_{0n'} \\ \xi_{10} & \cdots & \xi_{1n'} \\ \xi_{20} & \cdots & \xi_{2n'} \\ \xi_{30} & \cdots & \xi_{3n'} \\ \xi_{40} & \cdots & \xi_{4n'} \\ \vdots & & \vdots \\ \xi_{T0} & \cdots & \xi_{Tn'} \end{bmatrix}$$

CONDITIONAL PROBABILITY

$$P(y|x)P(x) = P(x|y)P(y)$$

BAYES RULE

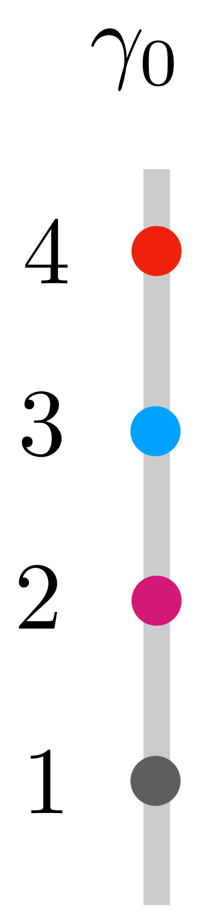
$$P(y|x) = \frac{P(x|y)P(y)}{P(x)}$$

$P(x|y=k)$ …height of density k

$P(y=k)$ …prior probability of k

$P(x)$ …normalize factor

$P(y=k|x)$ …probability of class k given x

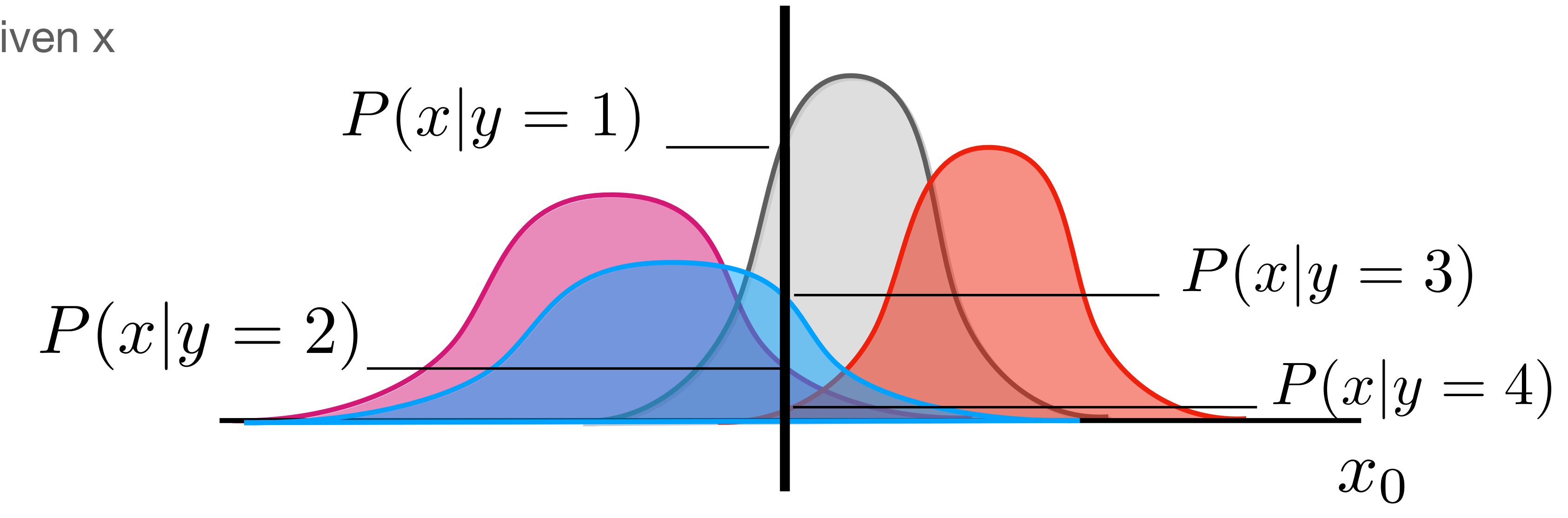


Class score

$$\log\ P(y=k|x) = -\tfrac{1}{2}\log|\Sigma_k| - \tfrac{1}{2}(x-\mu_k)^T\Sigma_k^{-1}(x-\mu_k) + \log P(y=k) + \text{const}$$

Predicted class

$$\arg\max_k \quad \log\ P(y=k|x)$$



$$P(x|y=k) = \frac{1}{(2\pi)^{n/2}|\Sigma_k|^{1/2}} e^{-\frac{1}{2}\left(x-\mu_k\right)^T\Sigma_k^{-1}(x-\mu_k)}$$

Discriminant Analysis

OUTPUTS
(Dependent Variables)

INPUTS
(Independent Variables)

$$\begin{bmatrix} \gamma_{00} & \cdots & \gamma_{0m'} & y_{00} & \cdots & y_{0m} \\ \gamma_{10} & \cdots & \gamma_{1m'} & y_{10} & \cdots & y_{1m} \\ \gamma_{20} & \cdots & \gamma_{2m'} & y_{20} & \cdots & y_{2m} \\ \gamma_{30} & \cdots & \gamma_{3m'} & y_{30} & \cdots & y_{3m} \\ \gamma_{40} & \cdots & \gamma_{4m'} & y_{40} & \cdots & y_{4m} \\ \vdots & & \vdots & \vdots & & \vdots \\ \gamma_{T0} & \cdots & \gamma_{Tm'} & y_{T0} & \cdots & y_{Tm} \end{bmatrix} \Leftarrow f\left[\begin{matrix} \Big(x_{00} & \cdots & x_{0n} \\ x_{10} & \cdots & x_{1n} \Big) \\ x_{20} & \cdots & x_{2n} \\ x_{30} & \cdots & x_{3n} \\ x_{40} & \cdots & x_{4n} \\ \vdots & & \vdots \\ x_{T0} & \cdots & x_{Tn} \end{matrix}\right] \begin{bmatrix} \xi_{00} & \cdots & \xi_{0n'} \\ \xi_{10} & \cdots & \xi_{1n'} \\ \xi_{20} & \cdots & \xi_{2n'} \\ \xi_{30} & \cdots & \xi_{3n'} \\ \xi_{40} & \cdots & \xi_{4n'} \\ \vdots & & \vdots \\ \xi_{T0} & \cdots & \xi_{Tn'} \end{bmatrix}$$

CONDITIONAL PROBABILITY

$$P(y|x)P(x) = P(x|y)P(y)$$

BAYES RULE

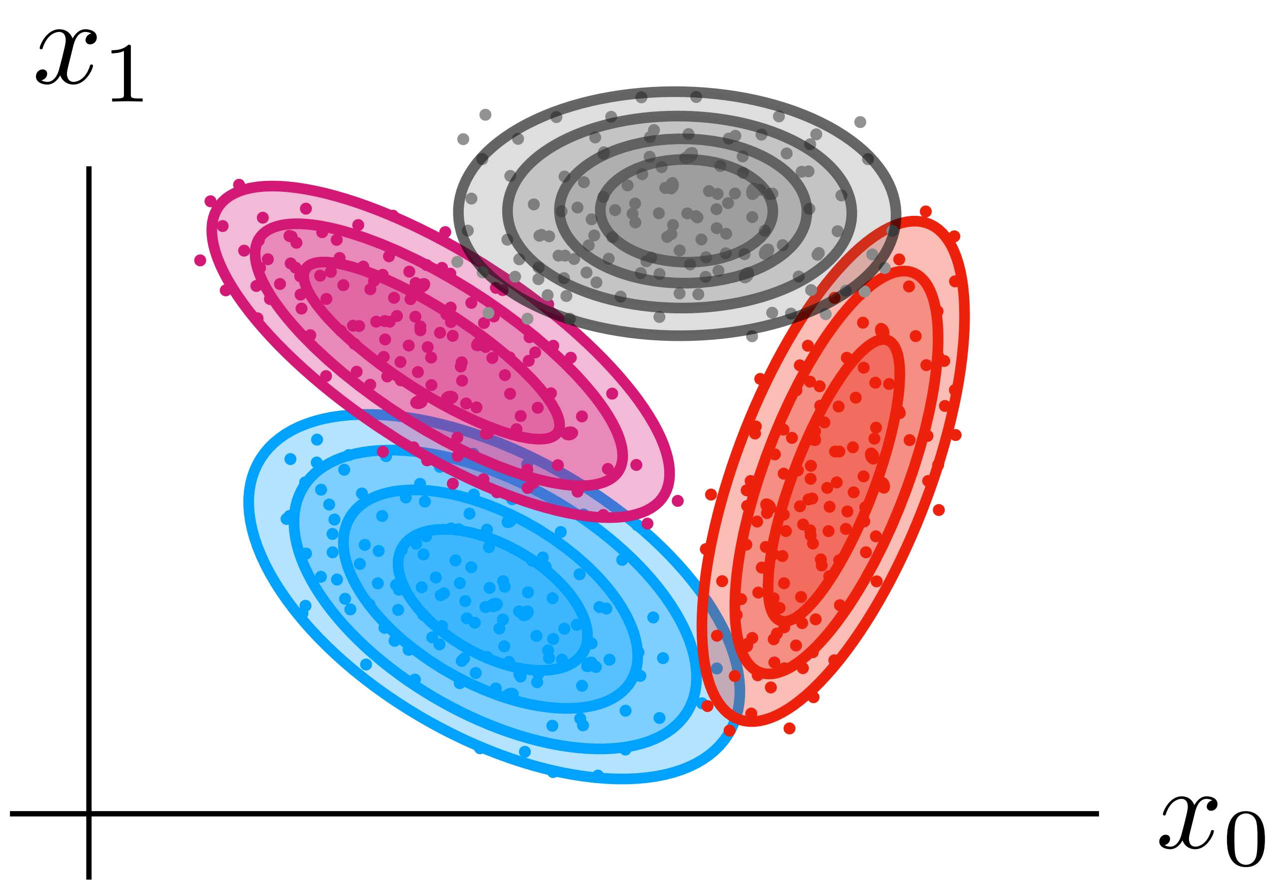
$$P(y|x) = \frac{P(x|y)P(y)}{P(x)}$$

$P(x|y=k)$ ...height of density k

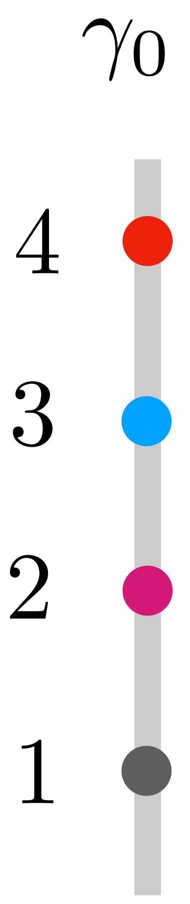
$P(y=k)$ ...prior probability of k

$P(x)$ ...normalize factor

$P(y=k|x)$ ...probability of class k given x



$$P(x|y=k) = \frac{1}{(2\pi)^{n/2}|\Sigma_k|^{1/2}} e^{-\frac{1}{2}(x-\mu_k)^T\Sigma_k^{-1}(x-\mu_k)}$$



Class score

$$\log\ P(y=k|x) = -\tfrac{1}{2}\log|\Sigma_k| - \tfrac{1}{2}(x-\mu_k)^T\Sigma_k^{-1}(x-\mu_k) + \log P(y=k) + \text{const}$$

Predicted class

$$\arg\max_k \ \log\ P(y=k|x)$$

Discrimination Surfaces

$$\log\ P(y=k|x) = \log\ P(y=k'|x)$$

k vs. k'

Quadratic Discriminant Analysis

$$\Rightarrow \boxed{\tfrac{1}{2}x^T\underbrace{\left(\Sigma_k^{-1} - \Sigma_{k'}^{-1}\right)}x - \left(\mu_k^T\Sigma_k^{-1} - \mu_{k'}^T\Sigma_{k'}^{-1}\right)x + C_{kk'}}$$

ellipses or hyperbolas (or parabolas)

with: $C_{kk'} = \mu_k^T\Sigma_k^{-1}\mu_k - \mu_{k'}^T\Sigma_{k'}^{-1}\mu_{k'} + \tfrac{1}{2}\log|\Sigma_k| - \tfrac{1}{2}\log|\Sigma_{k'}| + \log\ P(y=k') - \log\ P(y=k)$

Discriminant Analysis

OUTPUTS (Dependent Variables)

$$y_t = \theta^T x_t$$

INPUTS (Independent Variables)

$$\begin{bmatrix} \gamma_{00} & \cdots & \gamma_{0m'} & y_{00} & \cdots & y_{0m} \\ \gamma_{10} & \cdots & \gamma_{1m'} & y_{10} & \cdots & y_{1m} \\ \gamma_{20} & \cdots & \gamma_{2m'} & y_{20} & \cdots & y_{2m} \\ \gamma_{30} & \cdots & \gamma_{3m'} & y_{30} & \cdots & y_{3m} \\ \gamma_{40} & \cdots & \gamma_{4m'} & y_{40} & \cdots & y_{4m} \\ \vdots & & \vdots & \vdots & & \vdots \\ \gamma_{T0} & \cdots & \gamma_{Tm'} & y_{T0} & \cdots & y_{Tm} \end{bmatrix} \Leftarrow \boxed{f} \begin{bmatrix} \Big(x_{00} & \cdots & x_{0n} \Big) \\ x_{10} & \cdots & x_{1n} \\ x_{20} & \cdots & x_{2n} \\ x_{30} & \cdots & x_{3n} \\ x_{40} & \cdots & x_{4n} \\ \vdots & & \vdots \\ x_{T0} & \cdots & x_{Tn} \end{bmatrix} \begin{bmatrix} \xi_{00} & \cdots & \xi_{0n'} \\ \xi_{10} & \cdots & \xi_{1n'} \\ \xi_{20} & \cdots & \xi_{2n'} \\ \xi_{30} & \cdots & \xi_{3n'} \\ \xi_{40} & \cdots & \xi_{4n'} \\ \vdots & & \vdots \\ \xi_{T0} & \cdots & \xi_{Tn'} \end{bmatrix}$$

CONDITIONAL PROBABILITY

$$P(y|x)P(x) = P(x|y)P(y)$$

BAYES RULE

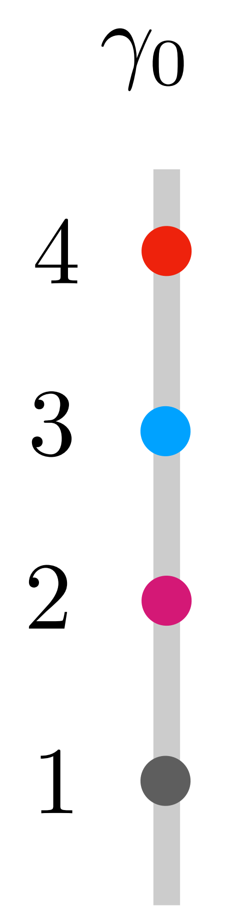
$$P(y|x) = \frac{P(x|y)P(y)}{P(x)}$$

$P(x|y=k)$ ...height of density k

$P(y=k)$ ...prior probability of k

$P(x)$ ...normalize factor

$P(y=k|x)$ ...probability of class k given x



Class score

$$\log\ P(y=k|x) = -\tfrac{1}{2}\log|\Sigma_k| - \tfrac{1}{2}(x-\mu_k)^T\Sigma_k^{-1}(x-\mu_k) + \log P(y=k) + \text{const}$$

Predicted class

$$\arg\max_k \quad \log\ P(y=k|x)$$

Discrimination Surfaces

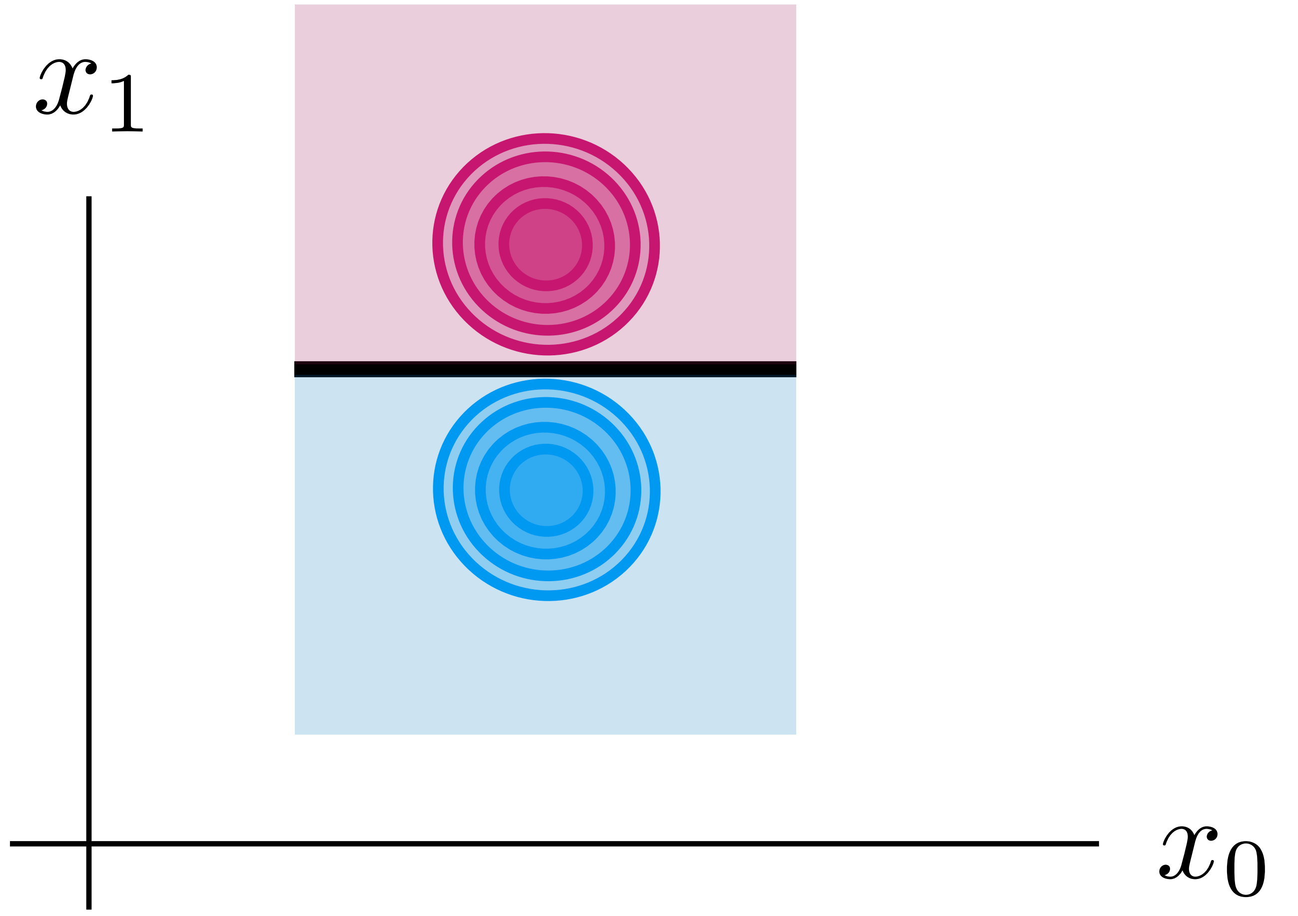
$$\log\ P(y=k|x) = \log\ P(y=k'|x)$$

k vs. k'

Quadratic Discriminant Analysis

$$\Rightarrow \boxed{\tfrac{1}{2}x^T\underbrace{\left(\Sigma_k^{-1} - \Sigma_{k'}^{-1}\right)}_{}x - \left(\mu_k^T\Sigma_k^{-1} - \mu_{k'}^T\Sigma_{k'}^{-1}\right)x + C_{kk'}}$$

ellipses or hyperbolas (or parabolas)



$$P(x|y=k) = \frac{1}{(2\pi)^{n/2}|\Sigma_k|^{1/2}} e^{-\frac{1}{2}(x-\mu_k)^T\Sigma_k^{-1}(x-\mu_k)}$$

with: $C_{kk'} = \mu_k^T\Sigma_k^{-1}\mu_k - \mu_{k'}^T\Sigma_{k'}^{-1}\mu_{k'} + \frac{1}{2}\log|\Sigma_k| - \frac{1}{2}\log|\Sigma_{k'}| + \log\ P(y=k') - \log\ P(y=k)$

Discriminant Analysis

OUTPUTS
(Dependent Variables)

$$y_t = \theta^T x_t$$

INPUTS
(Independent Variables)

CONDITIONAL PROBABILITY

$$P(y|x)P(x) = P(x|y)P(y)$$

BAYES RULE

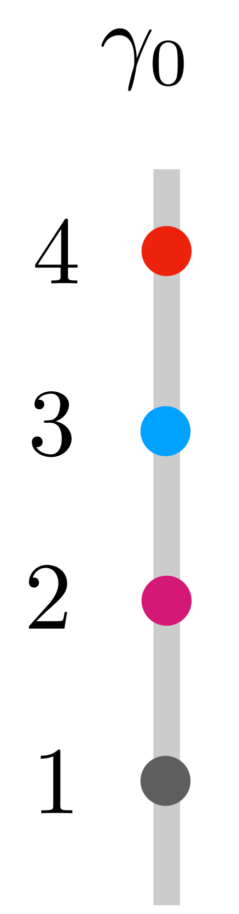
$$P(y|x) = \frac{P(x|y)P(y)}{P(x)}$$

$P(x|y=k)$ …height of density k

$P(y=k)$ …prior probability of k

$P(x)$ …normalize factor

$P(y=k|x)$ …probability of class k given x



Class score

$$\log\ P(y=k|x) = -\tfrac{1}{2}\log|\Sigma_k| - \tfrac{1}{2}(x-\mu_k)^T\Sigma_k^{-1}(x-\mu_k) + \log P(y=k) + \text{const}$$

Predicted class

$$\arg\max_k \quad \log\ P(y=k|x)$$

Discrimination Surfaces

$$\log\ P(y=k|x) = \log\ P(y=k'|x)$$

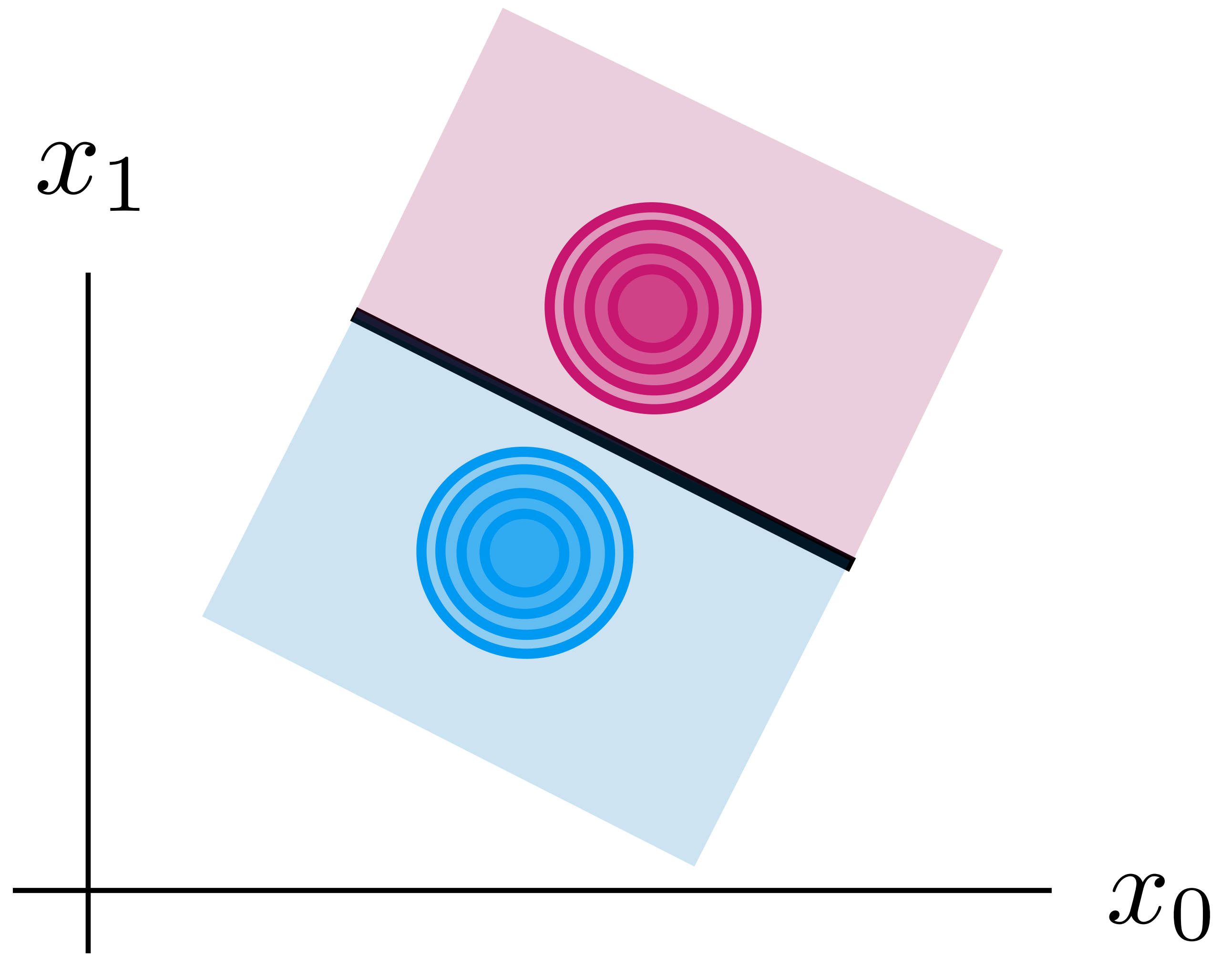
k vs. k'

Quadratic Discriminant Analysis

$$\Rightarrow \quad \boxed{\tfrac{1}{2}x^T\underbrace{\left(\Sigma_k^{-1} - \Sigma_{k'}^{-1}\right)}x - \left(\mu_k^T\Sigma_k^{-1} - \mu_{k'}^T\Sigma_{k'}^{-1}\right)x + C_{kk'}}$$

ellipses or hyperbolas (or parabolas)

with: $C_{kk'} = \mu_k^T\Sigma_k^{-1}\mu_k - \mu_{k'}^T\Sigma_{k'}^{-1}\mu_{k'} + \tfrac{1}{2}\log|\Sigma_k| - \tfrac{1}{2}\log|\Sigma_{k'}| + \log\ P(y=k') - \log\ P(y=k)$



$$P(x|y=k) = \frac{1}{(2\pi)^{n/2}|\Sigma_k|^{1/2}} e^{-\frac{1}{2}\left(x-\mu_k\right)^T\Sigma_k^{-1}\left(x-\mu_k\right)}$$

Discriminant Analysis

OUTPUTS
(Dependent Variables)

$$y_t = \theta^T x_t$$

INPUTS
(Independent Variables)

$$\begin{bmatrix} \gamma_{00} & \cdots & \gamma_{0m'} & y_{00} & \cdots & y_{0m} \\ \gamma_{10} & \cdots & \gamma_{1m'} & y_{10} & \cdots & y_{1m} \\ \gamma_{20} & \cdots & \gamma_{2m'} & y_{20} & \cdots & y_{2m} \\ \gamma_{30} & \cdots & \gamma_{3m'} & y_{30} & \cdots & y_{3m} \\ \gamma_{40} & \cdots & \gamma_{4m'} & y_{40} & \cdots & y_{4m} \\ \vdots & & \vdots & \vdots & & \vdots \\ \gamma_{T0} & \cdots & \gamma_{Tm'} & y_{T0} & \cdots & y_{Tm} \end{bmatrix} \Leftarrow \begin{bmatrix} f\left(\begin{matrix} x_{00} & \cdots & x_{0n} \\ x_{10} & \cdots & x_{1n} \end{matrix} \right) & & \xi_{00} & \cdots & \xi_{0n'} \\ & & \xi_{10} & \cdots & \xi_{1n'} \\ x_{20} & \cdots & x_{2n} & \xi_{20} & \cdots & \xi_{2n'} \\ x_{30} & \cdots & x_{3n} & \xi_{30} & \cdots & \xi_{3n'} \\ x_{40} & \cdots & x_{4n} & \xi_{40} & \cdots & \xi_{4n'} \\ \vdots & & \vdots & \vdots & & \vdots \\ x_{T0} & \cdots & x_{Tn} & \xi_{T0} & \cdots & \xi_{Tn'} \end{bmatrix}$$

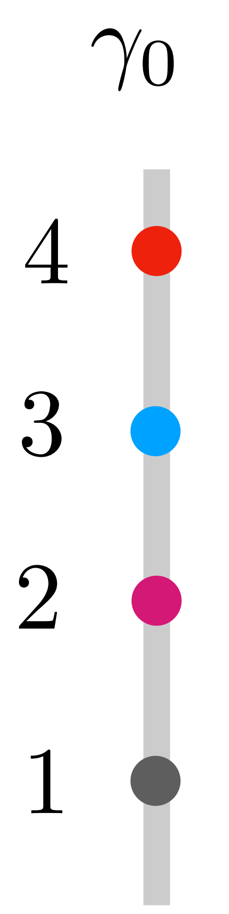
CONDITIONAL PROBABILITY

$$P(y|x)P(x) = P(x|y)P(y)$$

BAYES RULE

$$P(y|x) = \frac{P(x|y)P(y)}{P(x)}$$

$P(x|y=k)$ ...height of density k
$P(y=k)$ ...prior probability of k
$P(x)$ ...normalize factor
$P(y=k|x)$ ...probability of class k given x



Class score

$$\log\ P(y=k|x) = -\tfrac{1}{2}\log|\Sigma_k| - \tfrac{1}{2}(x-\mu_k)^T\Sigma_k^{-1}(x-\mu_k) + \log P(y=k) + \text{const}$$

Predicted class

$$\arg\max_k \quad \log\ P(y=k|x)$$

Discrimination Surfaces

$$\log\ P(y=k|x) = \log\ P(y=k'|x)$$

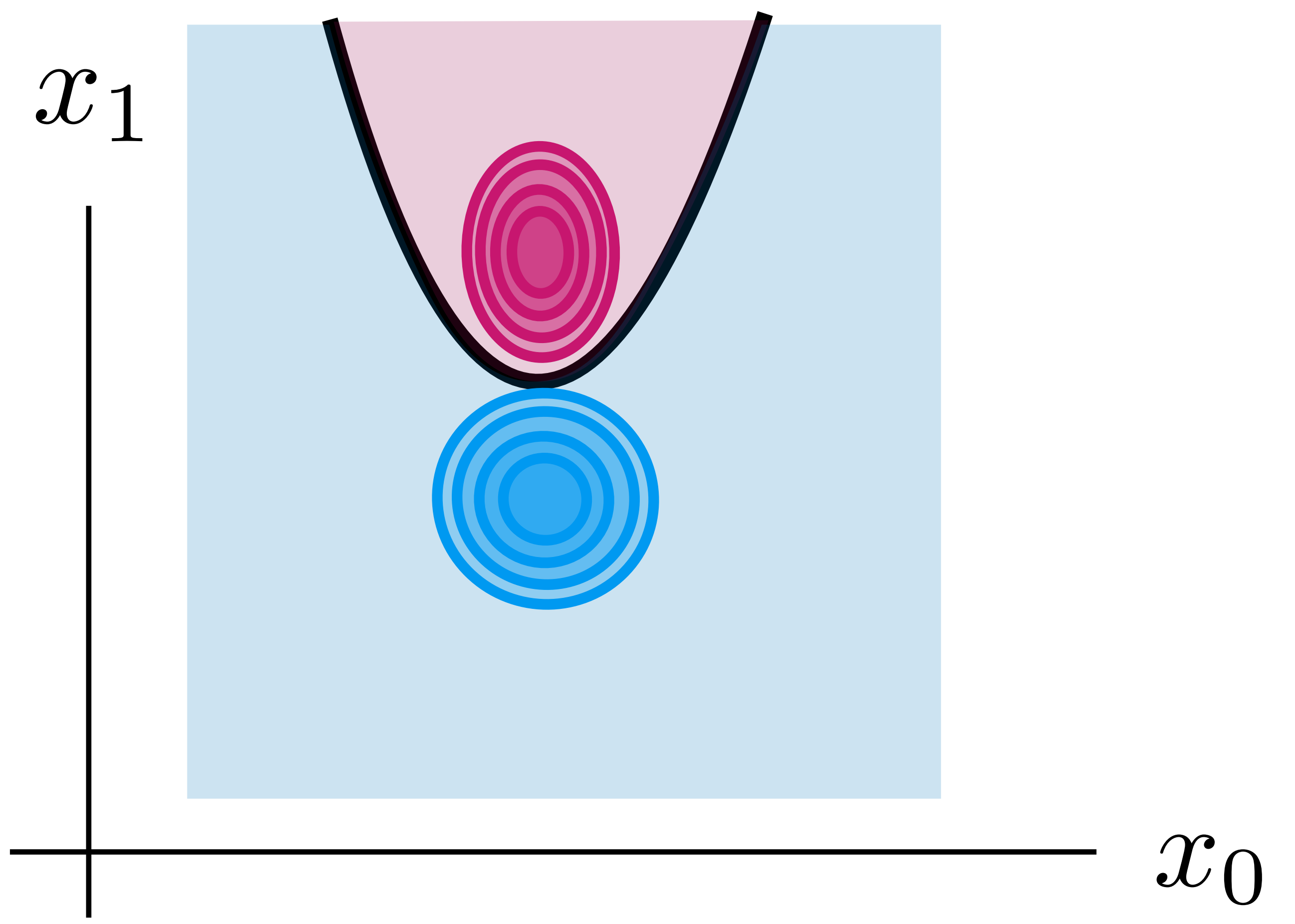
k vs. k'

Quadratic Discriminant Analysis

$$\Rightarrow \boxed{\tfrac{1}{2}x^T\underbrace{\left(\Sigma_k^{-1} - \Sigma_{k'}^{-1}\right)}_{}x - \left(\mu_k^T\Sigma_k^{-1} - \mu_{k'}^T\Sigma_{k'}^{-1}\right)x + C_{kk'}}$$

ellipses or hyperbolas (or parabolas)

with: $C_{kk'} = \mu_k^T\Sigma_k^{-1}\mu_k - \mu_{k'}^T\Sigma_{k'}^{-1}\mu_{k'} + \tfrac{1}{2}\log|\Sigma_k| - \tfrac{1}{2}\log|\Sigma_{k'}| + \log\ P(y=k') - \log\ P(y=k)$



$$P(x|y=k) = \frac{1}{(2\pi)^{n/2}|\Sigma_k|^{1/2}} e^{-\frac{1}{2}(x-\mu_k)^T\Sigma_k^{-1}(x-\mu_k)}$$

Discriminant Analysis

OUTPUTS

(Dependent Variables)

$$y_t = \theta^T x_t$$

INPUTS

(Independent Variables)

$$\begin{bmatrix} \gamma_{00} & \cdots & \gamma_{0m'} & y_{00} & \cdots & y_{0m} \\ \gamma_{10} & \cdots & \gamma_{1m'} & y_{10} & \cdots & y_{1m} \\ \gamma_{20} & \cdots & \gamma_{2m'} & y_{20} & \cdots & y_{2m} \\ \gamma_{30} & \cdots & \gamma_{3m'} & y_{30} & \cdots & y_{3m} \\ \gamma_{40} & \cdots & \gamma_{4m'} & y_{40} & \cdots & y_{4m} \\ \vdots & & \vdots & \vdots & & \vdots \\ \gamma_{T0} & \cdots & \gamma_{Tm'} & y_{T0} & \cdots & y_{Tm} \end{bmatrix} \Leftarrow \boxed{f\left[\begin{pmatrix} x_{00} & \cdots & x_{0n} \\ x_{10} & \cdots & x_{1n} \end{pmatrix}\right.} \begin{matrix} \xi_{00} & \cdots & \xi_{0n'} \\ \xi_{10} & \cdots & \xi_{1n'} \end{matrix}$$

$$\left.\begin{matrix} x_{20} & \cdots & x_{2n} \\ x_{30} & \cdots & x_{3n} \\ x_{40} & \cdots & x_{4n} \\ \vdots & & \vdots \\ x_{T0} & \cdots & x_{Tn} \end{matrix}\right] \begin{matrix} \xi_{20} & \cdots & \xi_{2n'} \\ \xi_{30} & \cdots & \xi_{3n'} \\ \xi_{40} & \cdots & \xi_{4n'} \\ \vdots & & \vdots \\ \xi_{T0} & \cdots & \xi_{Tn'} \end{matrix}$$

CONDITIONAL PROBABILITY

$$P(y|x)P(x) = P(x|y)P(y)$$

BAYES RULE

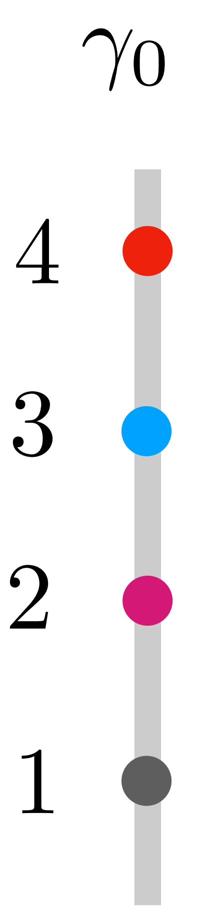
$$P(y|x) = \frac{P(x|y)P(y)}{P(x)}$$

$P(x|y=k)$ ...height of density k

$P(y=k)$ ...prior probability of k

$P(x)$ ...normalize factor

$P(y=k|x)$ ...probability of class k given x



Class score

$$\log\ P(y=k|x) = -\tfrac{1}{2}\log|\Sigma_k| - \tfrac{1}{2}(x-\mu_k)^T\Sigma_k^{-1}(x-\mu_k) + \log P(y=k) + \text{const}$$

Predicted class

$$\arg\max_k \quad \log\ P(y=k|x)$$

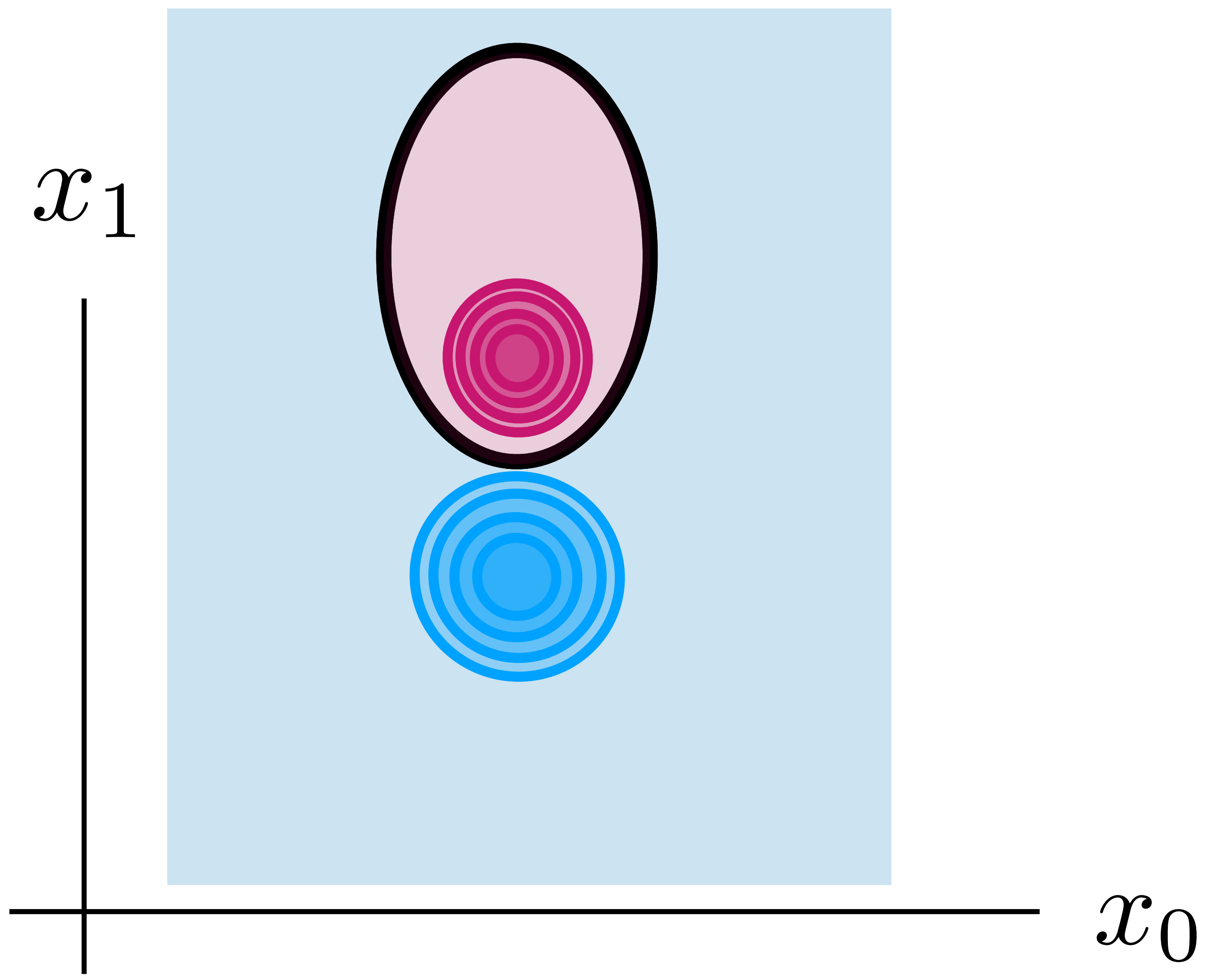
Discrimination Surfaces

$$\log\ P(y=k|x) = \log\ P(y=k'|x) \qquad \text{k vs. k'}$$

Quadratic Discriminant Analysis

$$\Rightarrow \boxed{\tfrac{1}{2}x^T\underbrace{\left(\Sigma_k^{-1} - \Sigma_{k'}^{-1}\right)}_{}x - \left(\mu_k^T\Sigma_k^{-1} - \mu_{k'}^T\Sigma_{k'}^{-1}\right)x + C_{kk'}}$$

ellipses or hyperbolas (or parabolas)



$$P(x|y=k) = \frac{1}{(2\pi)^{n/2}|\Sigma_k|^{1/2}} e^{-\frac{1}{2}(x-\mu_k)^T\Sigma_k^{-1}(x-\mu_k)}$$

with: $C_{kk'} = \mu_k^T\Sigma_k^{-1}\mu_k - \mu_{k'}^T\Sigma_{k'}^{-1}\mu_{k'} + \tfrac{1}{2}\log|\Sigma_k| - \tfrac{1}{2}\log|\Sigma_{k'}| + \log\ P(y=k') - \log\ P(y=k)$

Discriminant Analysis

OUTPUTS (Dependent Variables)

$$y_t = \theta^T x_t$$

INPUTS (Independent Variables)

$$\begin{bmatrix} \gamma_{00} & \cdots & \gamma_{0m'} & y_{00} & \cdots & y_{0m} \\ \gamma_{10} & \cdots & \gamma_{1m'} & y_{10} & \cdots & y_{1m} \\ \gamma_{20} & \cdots & \gamma_{2m'} & y_{20} & \cdots & y_{2m} \\ \gamma_{30} & \cdots & \gamma_{3m'} & y_{30} & \cdots & y_{3m} \\ \gamma_{40} & \cdots & \gamma_{4m'} & y_{40} & \cdots & y_{4m} \\ \vdots & & \vdots & \vdots & & \vdots \\ \gamma_{T0} & \cdots & \gamma_{Tm'} & y_{T0} & \cdots & y_{Tm} \end{bmatrix} \Leftarrow f\begin{bmatrix} \begin{pmatrix} x_{00} & \cdots & x_{0n} \\ x_{10} & \cdots & x_{1n} \end{pmatrix} \\ \begin{matrix} x_{20} & \cdots & x_{2n} \\ x_{30} & \cdots & x_{3n} \\ x_{40} & \cdots & x_{4n} \\ \vdots & & \vdots \\ x_{T0} & \cdots & x_{Tn} \end{matrix} \end{bmatrix} \begin{bmatrix} \xi_{00} & \cdots & \xi_{0n'} \\ \xi_{10} & \cdots & \xi_{1n'} \\ \xi_{20} & \cdots & \xi_{2n'} \\ \xi_{30} & \cdots & \xi_{3n'} \\ \xi_{40} & \cdots & \xi_{4n'} \\ \vdots & & \vdots \\ \xi_{T0} & \cdots & \xi_{Tn'} \end{bmatrix}$$

CONDITIONAL PROBABILITY

$$P(y|x)P(x) = P(x|y)P(y)$$

BAYES RULE

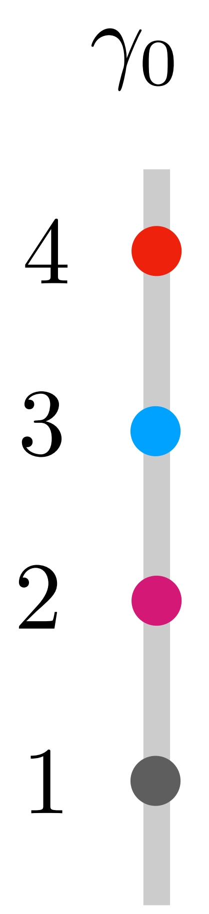
$$P(y|x) = \frac{P(x|y)P(y)}{P(x)}$$

$P(x|y=k)$ ...height of density k

$P(y=k)$ ...prior probability of k

$P(x)$ ...normalize factor

$P(y=k|x)$ ...probability of class k given x



Class score

$$\log\ P(y=k|x) = -\tfrac{1}{2}\log|\Sigma_k| - \tfrac{1}{2}(x-\mu_k)^T\Sigma_k^{-1}(x-\mu_k) + \log P(y=k) + \text{const}$$

Predicted class

$$\arg\max_k \quad \log\ P(y=k|x)$$

Discrimination Surfaces

$$\log\ P(y=k|x) = \log\ P(y=k'|x)$$

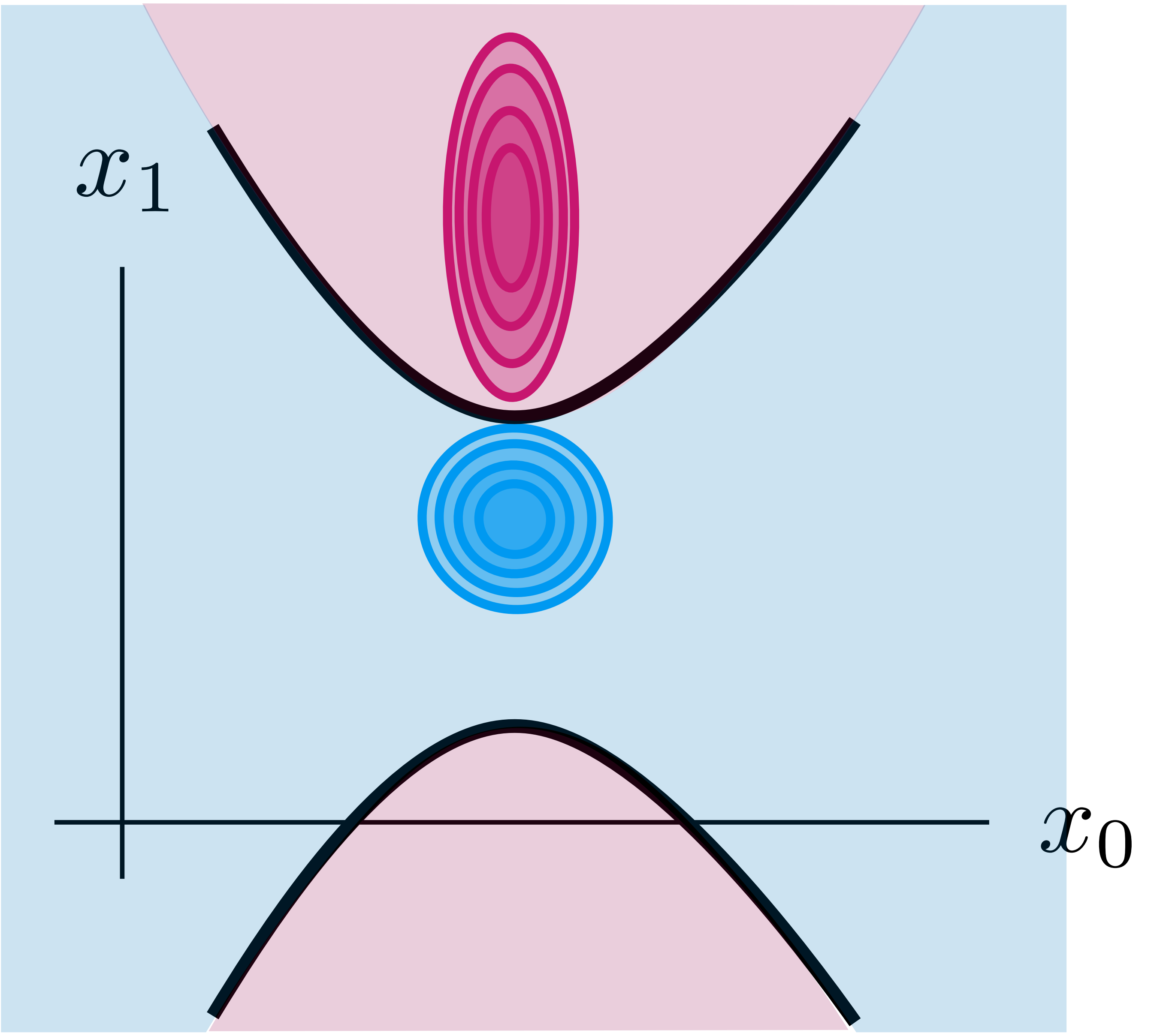
k vs. k'

Quadratic Discriminant Analysis

$$\Rightarrow \boxed{\tfrac{1}{2}x^T\underbrace{\left(\Sigma_k^{-1} - \Sigma_{k'}^{-1}\right)}_{}x - \left(\mu_k^T\Sigma_k^{-1} - \mu_{k'}^T\Sigma_{k'}^{-1}\right)x + C_{kk'}}$$

ellipses or hyperbolas (or parabolas)

with: $C_{kk'} = \mu_k^T\Sigma_k^{-1}\mu_k - \mu_{k'}^T\Sigma_{k'}^{-1}\mu_{k'} + \tfrac{1}{2}\log|\Sigma_k| - \tfrac{1}{2}\log|\Sigma_{k'}| + \log\ P(y=k') - \log\ P(y=k)$



$$P(x|y=k) = \frac{1}{(2\pi)^{n/2}|\Sigma_k|^{1/2}} e^{-\frac{1}{2}(x-\mu_k)^T\Sigma_k^{-1}(x-\mu_k)}$$

Discriminant Analysis

OUTPUTS

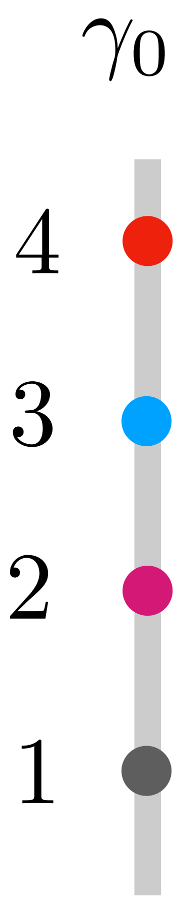
(Dependent Variables)

$$y_t = \theta^T x_t$$

INPUTS

(Independent Variables)

$$\left[\begin{array}{ccc|ccc} \boxed{\begin{matrix}\gamma_{00}\\ \gamma_{10}\\ \gamma_{20}\\ \gamma_{30}\\ \gamma_{40}\\ \vdots\\ \gamma_{T0}\end{matrix}} & \begin{matrix}\cdots\\ \cdots\\ \cdots\\ \cdots\\ \cdots\\ \\ \cdots\end{matrix} & \begin{matrix}\gamma_{0m'}\\ \gamma_{1m'}\\ \gamma_{2m'}\\ \gamma_{3m'}\\ \gamma_{4m'}\\ \vdots\\ \gamma_{Tm'}\end{matrix} & \begin{matrix}y_{00}\\ y_{10}\\ y_{20}\\ y_{30}\\ y_{40}\\ \vdots\\ y_{T0}\end{matrix} & \begin{matrix}\cdots\\ \cdots\\ \cdots\\ \cdots\\ \cdots\\ \\ \cdots\end{matrix} & \begin{matrix}y_{0m}\\ y_{1m}\\ y_{2m}\\ y_{3m}\\ y_{4m}\\ \vdots\\ y_{Tm}\end{matrix} \end{array}\right] \Leftarrow \boxed{f\left[\left(\begin{matrix} x_{00} & \cdots & x_{0n}\\ x_{10} & \cdots & x_{1n}\end{matrix}\right)\right.} \begin{matrix} \\ \\ x_{20} & \cdots & x_{2n}\\ x_{30} & \cdots & x_{3n}\\ x_{40} & \cdots & x_{4n}\\ \vdots & & \vdots\\ x_{T0} & \cdots & x_{Tn}\end{matrix}\Bigg] \begin{bmatrix} \xi_{00} & \cdots & \xi_{0n'}\\ \xi_{10} & \cdots & \xi_{1n'}\\ \xi_{20} & \cdots & \xi_{2n'}\\ \xi_{30} & \cdots & \xi_{3n'}\\ \xi_{40} & \cdots & \xi_{4n'}\\ \vdots & & \vdots\\ \xi_{T0} & \cdots & \xi_{Tn'}\end{bmatrix}$$



CONDITIONAL PROBABILITY

$$P(y|x)P(x) = P(x|y)P(y)$$

BAYES RULE

$$P(y|x) = \frac{P(x|y)P(y)}{P(x)}$$

$P(x|y=k)$ ...height of density k

$P(y=k)$ ...prior probability of k

$P(x)$ ...normalize factor

$P(y=k|x)$ ...probability of class k given x

Class score

$$\log\ P(y=k|x) = -\tfrac{1}{2}\log|\Sigma_k| - \tfrac{1}{2}(x-\mu_k)^T\Sigma_k^{-1}(x-\mu_k) + \log P(y=k) + \text{const}$$

Predicted class

$$\arg\max_k \quad \log\ P(y=k|x)$$

Discrimination Surfaces

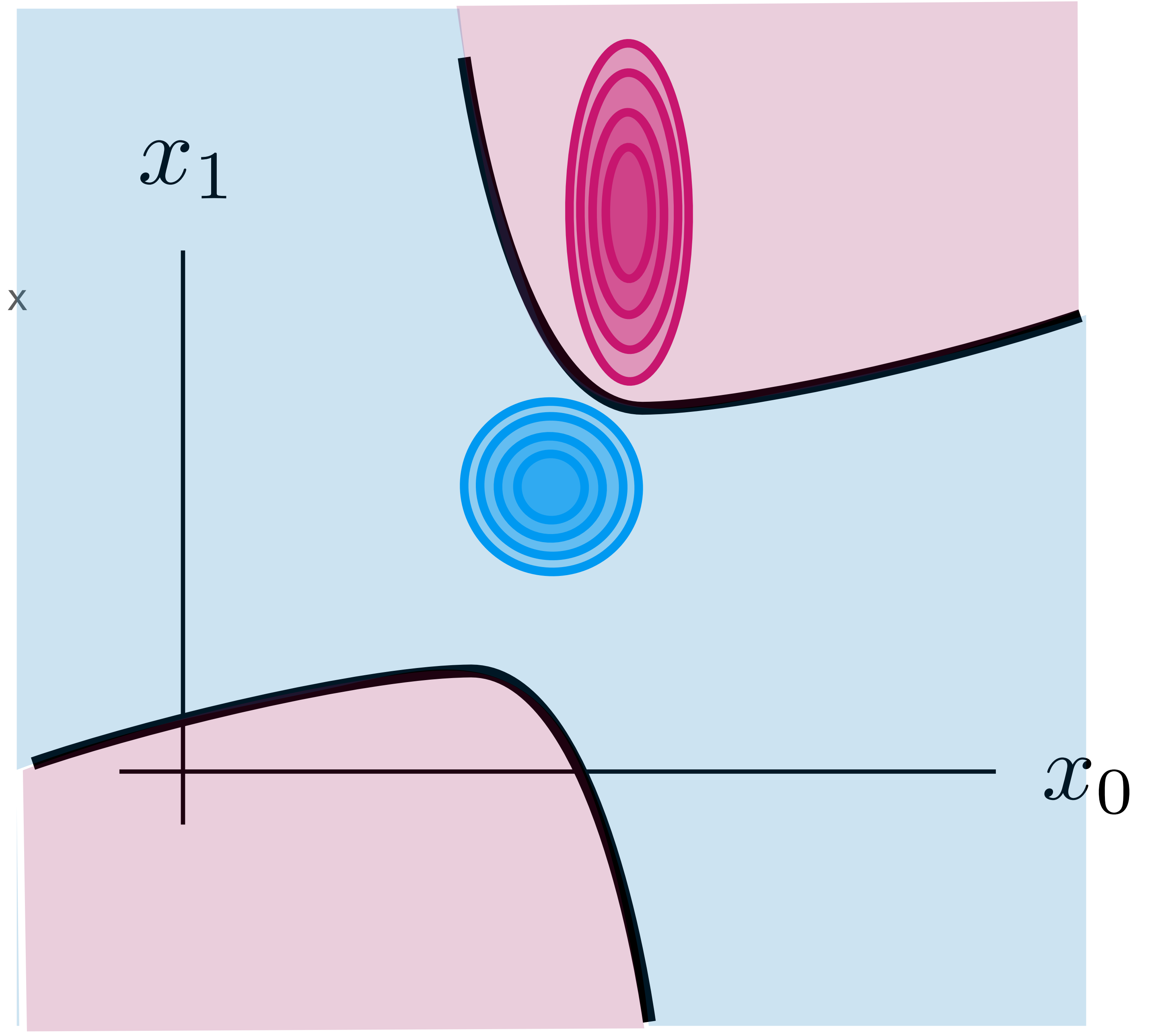
$$\log\ P(y=k|x) = \log\ P(y=k'|x)$$

k vs. k'

Quadratic Discriminant Analysis

$$\Rightarrow \boxed{\tfrac{1}{2}x^T\underbrace{\left(\Sigma_k^{-1} - \Sigma_{k'}^{-1}\right)}x - \left(\mu_k^T\Sigma_k^{-1} - \mu_{k'}^T\Sigma_{k'}^{-1}\right)x + C_{kk'}}$$

ellipses or hyperbolas (or parabolas)



$$P(x|y=k) = \frac{1}{(2\pi)^{n/2}|\Sigma_k|^{1/2}} e^{-\frac{1}{2}\left(x-\mu_k\right)^T\Sigma_k^{-1}\left(x-\mu_k\right)}$$

with: $C_{kk'} = \mu_k^T\Sigma_k^{-1}\mu_k - \mu_{k'}^T\Sigma_{k'}^{-1}\mu_{k'} + \tfrac{1}{2}\log|\Sigma_k| - \tfrac{1}{2}\log|\Sigma_{k'}| + \log\ P(y=k') - \log\ P(y=k)$

Discriminant Analysis

OUTPUTS (Dependent Variables)

$$y_t = \theta^T x_t$$

INPUTS (Independent Variables)

$$\begin{bmatrix} \boxed{\begin{matrix}\gamma_{00}\\ \gamma_{10}\\ \gamma_{20}\\ \gamma_{30}\\ \gamma_{40}\\ \vdots \\ \gamma_{T0}\end{matrix}} & \begin{matrix}\cdots\\ \cdots\\ \cdots\\ \cdots\\ \cdots\\ \\ \cdots\end{matrix} & \begin{matrix}\gamma_{0m'}\\ \gamma_{1m'}\\ \gamma_{2m'}\\ \gamma_{3m'}\\ \gamma_{4m'}\\ \vdots \\ \gamma_{Tm'}\end{matrix} & \begin{matrix}y_{00}\\ y_{10}\\ y_{20}\\ y_{30}\\ y_{40}\\ \vdots \\ y_{T0}\end{matrix} & \begin{matrix}\cdots\\ \cdots\\ \cdots\\ \cdots\\ \cdots\\ \\ \cdots\end{matrix} & \begin{matrix}y_{0m}\\ y_{1m}\\ y_{2m}\\ y_{3m}\\ y_{4m}\\ \vdots \\ y_{Tm}\end{matrix} \end{bmatrix} \Leftarrow \begin{bmatrix} \boxed{f\left(\begin{matrix} x_{00} & \cdots & x_{0n} \\ x_{10} & \cdots & x_{1n}\end{matrix}\right)} \\ \begin{matrix} x_{20} & \cdots & x_{2n} \\ x_{30} & \cdots & x_{3n} \\ x_{40} & \cdots & x_{4n} \\ \vdots & & \vdots \\ x_{T0} & \cdots & x_{Tn} \end{matrix} \end{bmatrix} \begin{bmatrix} \xi_{00} & \cdots & \xi_{0n'} \\ \xi_{10} & \cdots & \xi_{1n'} \\ \xi_{20} & \cdots & \xi_{2n'} \\ \xi_{30} & \cdots & \xi_{3n'} \\ \xi_{40} & \cdots & \xi_{4n'} \\ \vdots & & \vdots \\ \xi_{T0} & \cdots & \xi_{Tn'} \end{bmatrix}$$

CONDITIONAL PROBABILITY

$$P(y|x)P(x) = P(x|y)P(y)$$

BAYES RULE

$$P(y|x) = \frac{P(x|y)P(y)}{P(x)}$$

$P(x|y=k)$ ...height of density k

$P(y=k)$ ...prior probability of k

$P(x)$ ...normalize factor

$P(y=k|x)$ ...probability of class k given x

γ_0

4, 3, 2, 1

Class score

$$\log\ P(y=k|x) = -\tfrac{1}{2}\log|\Sigma_k| - \tfrac{1}{2}(x-\mu_k)^T\Sigma_k^{-1}(x-\mu_k) + \log P(y=k) + \text{const}$$

Predicted class

$$\arg\max_k \ \log\ P(y=k|x)$$

Discrimination Surfaces

$$\log\ P(y=k|x) = \log\ P(y=k'|x)$$

k vs. k'

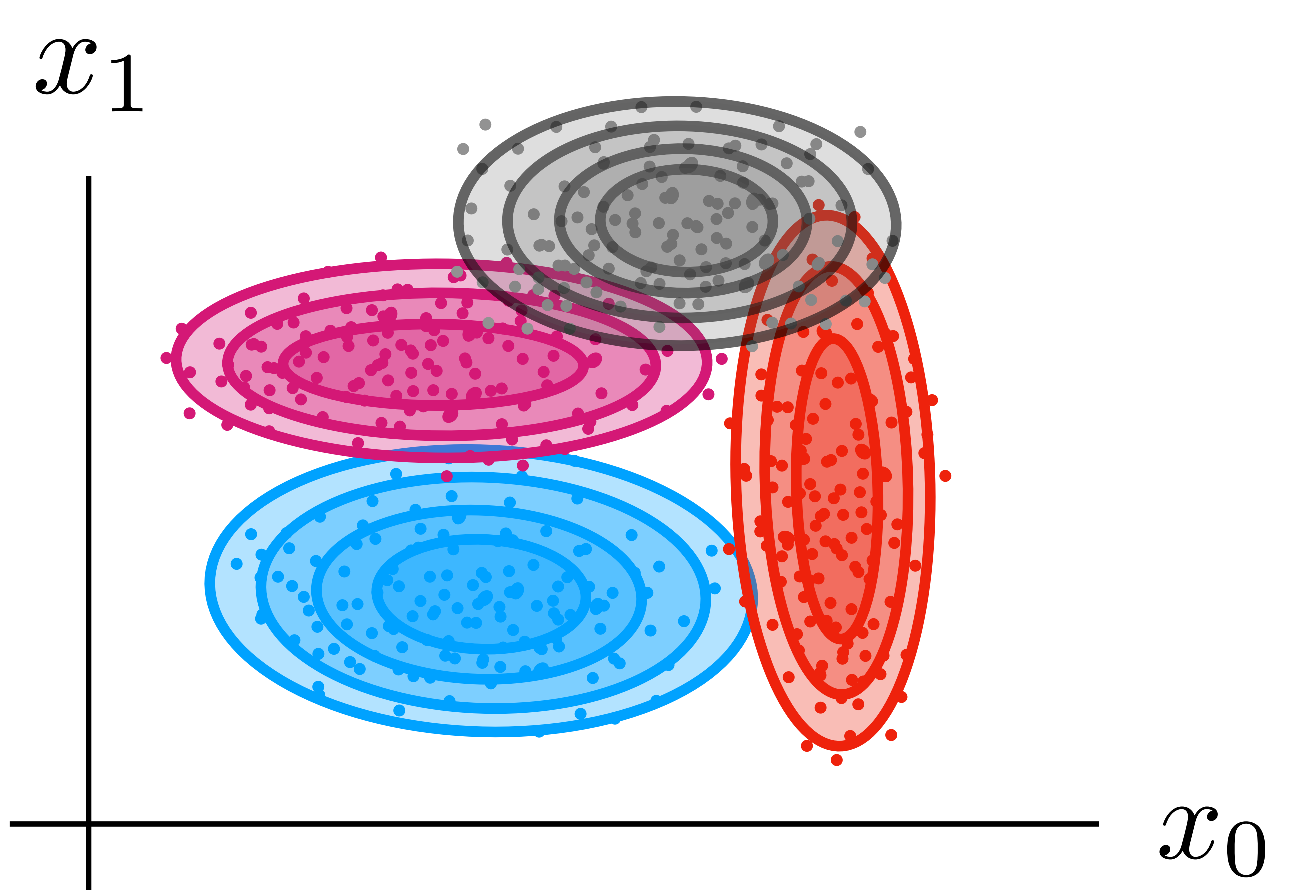
$$\Rightarrow \boxed{\tfrac{1}{2}x^T\underbrace{(\Sigma_k^{-1} - \Sigma_{k'}^{-1})}_{}x - (\mu_k^T\Sigma_k^{-1} - \mu_{k'}^T\Sigma_{k'}^{-1})x + C_{kk'}}$$

ellipses or hyperbolas (or parabolas)

Gaussian Naive Bayes Classifier

$\Sigma_k, \Sigma_{k'}$ **diagonal**

with: $C_{kk'} = \mu_k^T\Sigma_k^{-1}\mu_k - \mu_{k'}^T\Sigma_{k'}^{-1}\mu_{k'} + \tfrac{1}{2}\log|\Sigma_k| - \tfrac{1}{2}\log|\Sigma_{k'}| + \log\ P(y=k') - \log\ P(y=k)$



$$P(x|y=k) = \frac{1}{(2\pi)^{n/2}|\Sigma_k|^{1/2}} e^{-\frac{1}{2}(x-\mu_k)^T\Sigma_k^{-1}(x-\mu_k)}$$

Discriminant Analysis

OUTPUTS (Dependent Variables)

$$y_t = \theta^T x_t$$

INPUTS (Independent Variables)

$$\left[\begin{array}{ccc|ccc} \gamma_{00} & \cdots & \gamma_{0m'} & y_{00} & \cdots & y_{0m} \\ \gamma_{10} & \cdots & \gamma_{1m'} & y_{10} & \cdots & y_{1m} \\ \gamma_{20} & \cdots & \gamma_{2m'} & y_{20} & \cdots & y_{2m} \\ \gamma_{30} & \cdots & \gamma_{3m'} & y_{30} & \cdots & y_{3m} \\ \gamma_{40} & \cdots & \gamma_{4m'} & y_{40} & \cdots & y_{4m} \\ \vdots & & \vdots & \vdots & & \vdots \\ \gamma_{T0} & \cdots & \gamma_{Tm'} & y_{T0} & \cdots & y_{Tm} \end{array}\right] \Leftarrow \boxed{f} \left[\begin{array}{ccc|ccc} \boxed{(x_{00} & \cdots & x_{0n})} & \xi_{00} & \cdots & \xi_{0n'} \\ x_{10} & \cdots & x_{1n} & \xi_{10} & \cdots & \xi_{1n'} \\ x_{20} & \cdots & x_{2n} & \xi_{20} & \cdots & \xi_{2n'} \\ x_{30} & \cdots & x_{3n} & \xi_{30} & \cdots & \xi_{3n'} \\ x_{40} & \cdots & x_{4n} & \xi_{40} & \cdots & \xi_{4n'} \\ \vdots & & \vdots & \vdots & & \vdots \\ x_{T0} & \cdots & x_{Tn} & \xi_{T0} & \cdots & \xi_{Tn'} \end{array}\right]$$

CONDITIONAL PROBABILITY

$$P(y|x)P(x) = P(x|y)P(y)$$

BAYES RULE

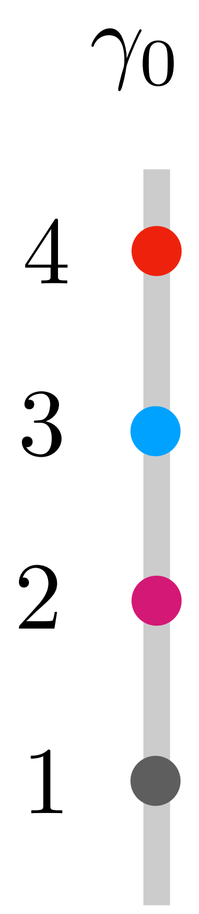
$$P(y|x) = \frac{P(x|y)P(y)}{P(x)}$$

$P(x|y=k)$ ...height of density k

$P(y=k)$ ...prior probability of k

$P(x)$ ...normalize factor

$P(y=k|x)$ ...probability of class k given x



Class score

$$\log\ P(y=k|x) = -\tfrac{1}{2}\log|\Sigma_k| - \boxed{\underbrace{\tfrac{1}{2}(x-\mu_k)^T\Sigma_k^{-1}(x-\mu_k)}_{\text{Mahalanobis Distance}}} + \log P(y=k) + \text{const}$$

Predicted class

$$\underset{k}{\arg\max}\ \ \log\ P(y=k|x)$$

Discrimination Surfaces

$$\log\ P(y=k|x) = \log\ P(y=k'|x) \qquad \text{k vs. k'}$$

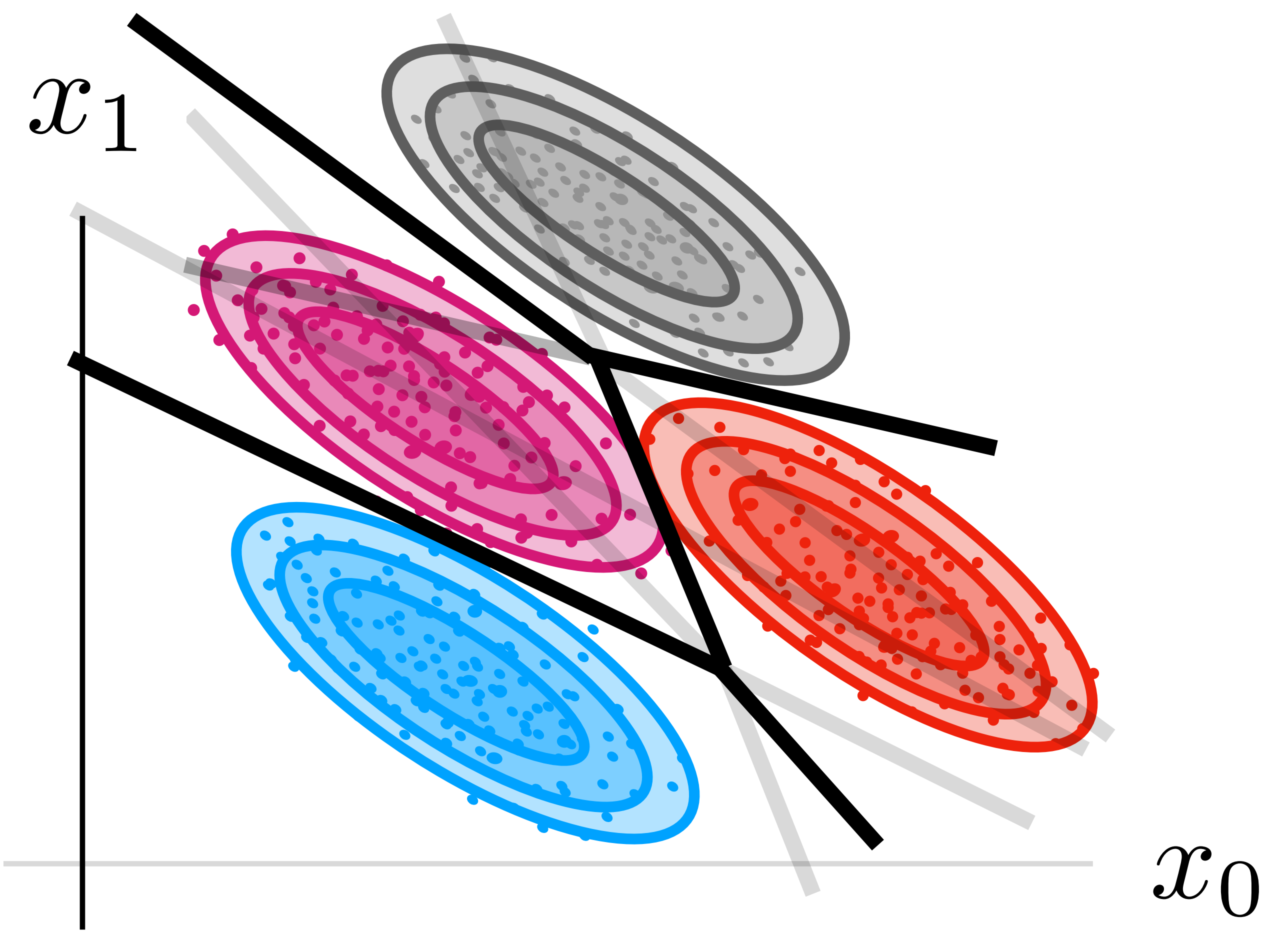
Linear Discriminant Analysis

$$\Rightarrow \boxed{\tfrac{1}{2}x^T\left(\cancelto{0}{\Sigma_k^{-1} - \Sigma_{k'}^{-1}}\right)x - \underbrace{\left(\mu_k^T\Sigma_k^{-1} - \mu_{k'}^T\Sigma_{k'}^{-1}\right)}x + C_{kk'}}$$

$$\Rightarrow \textbf{ hyperplanes}$$

$$\Sigma_k = \Sigma_{k'}$$

with: $C_{kk'} = \mu_k^T\Sigma_k^{-1}\mu_k - \mu_{k'}^T\Sigma_{k'}^{-1}\mu_{k'} + \tfrac{1}{2}\log|\Sigma_k| - \tfrac{1}{2}\log|\Sigma_{k'}| + \log\ P(y=k') - \log\ P(y=k)$



$$P(x|y=k) = \frac{1}{(2\pi)^{n/2}|\Sigma_k|^{1/2}} e^{-\frac{1}{2}(x-\mu_k)^T\Sigma_k^{-1}(x-\mu_k)}$$