Classification

ML - Supervised Learning

Dan Calderone - Win22

Classification

OUTPUTS

(Dependent Variables)

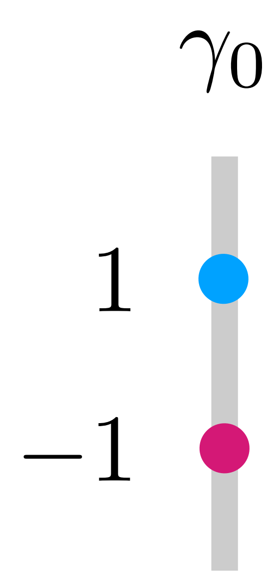
$$y_t = \theta^T x_t$$

INPUTS

(Independent Variables)

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

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



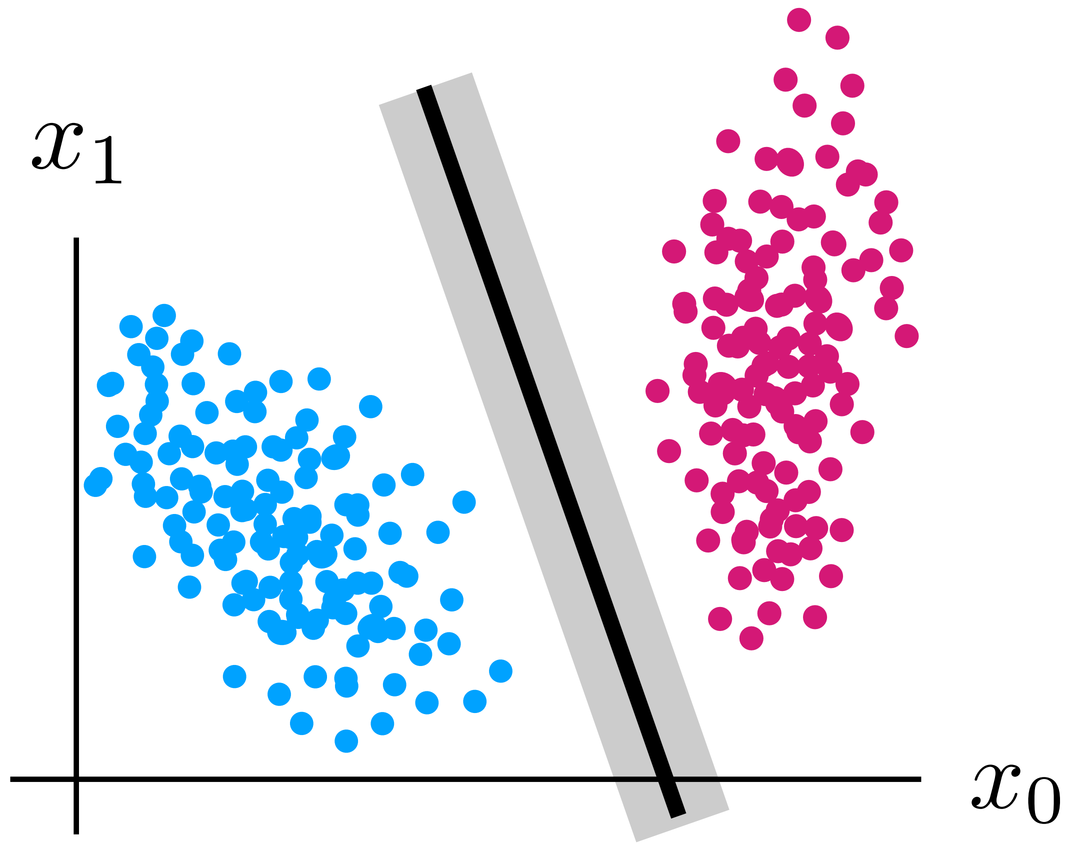
COST:

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

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

Support Vector Machine (SVM)

Hard boundary



"Binary classifier"

Classification

OUTPUTS

(Dependent Variables)

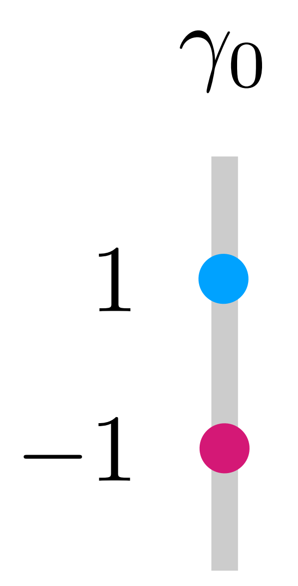
$$y_t = \theta^T x_t$$

INPUTS

(Independent Variables)

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

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

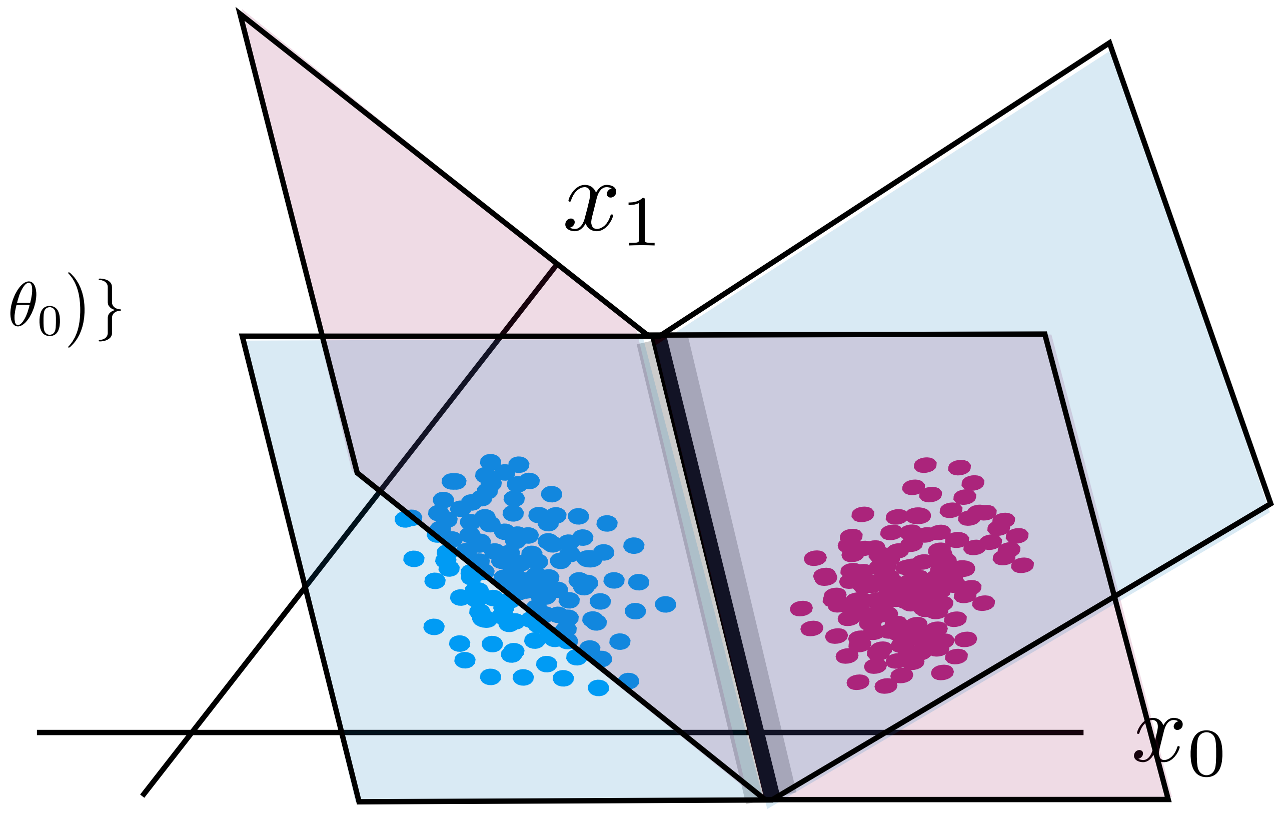


COST:

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

Support Vector Machine (SVM)

Soft boundary



"Binary classifier"

Classification

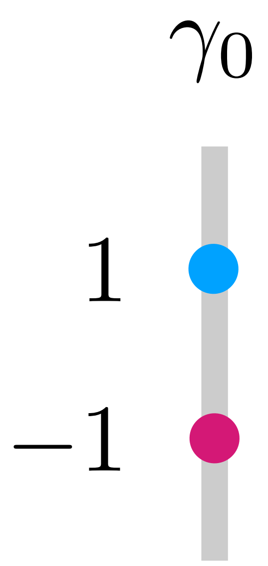
OUTPUTS
(Dependent Variables)

$$y_t = \theta^T x_t$$

INPUTS
(Independent Variables)

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

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

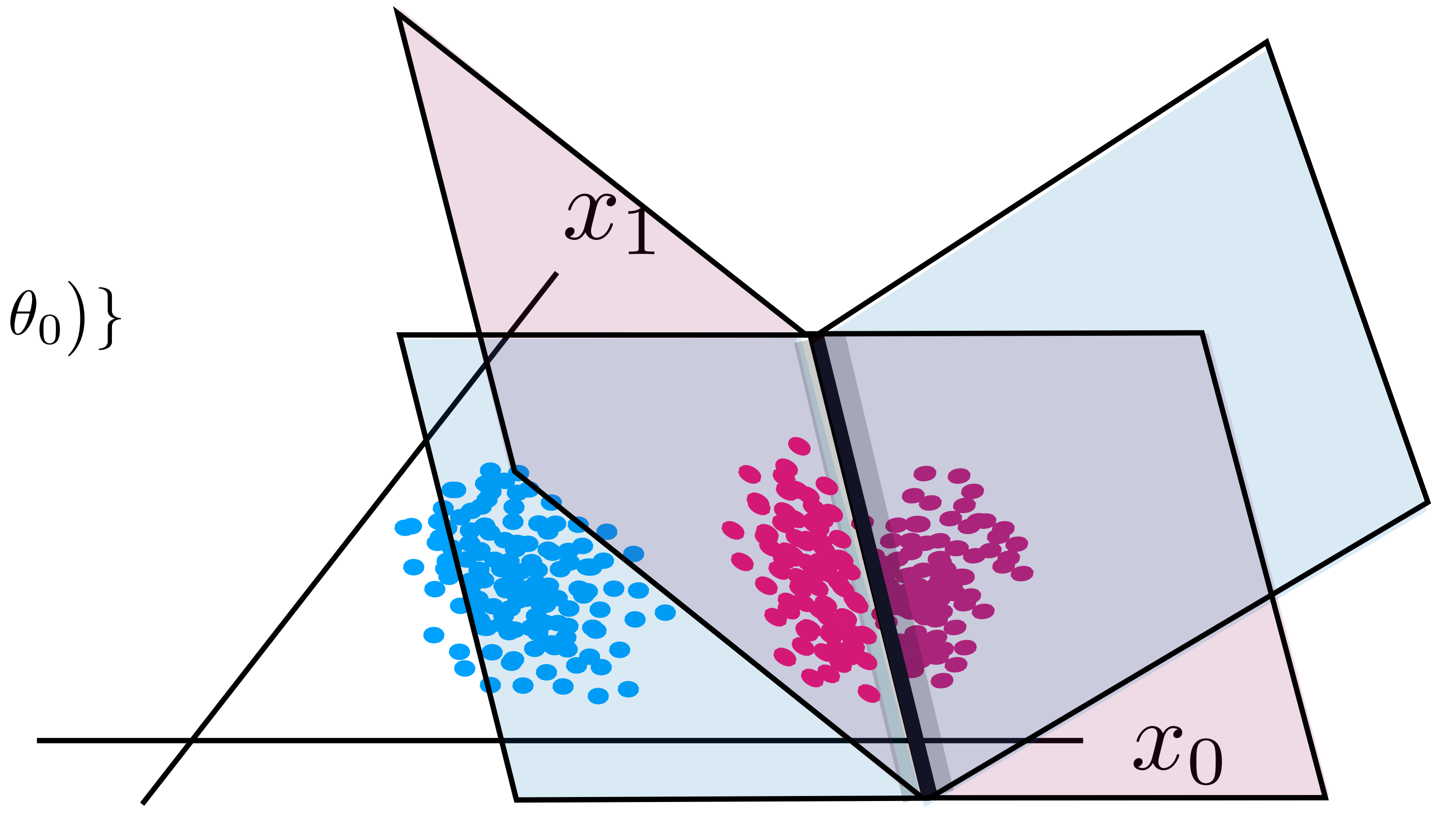


COST:

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

Support Vector Machine (SVM)

Soft boundary



"Binary classifier"

Classification

OUTPUTS

(Dependent Variables)

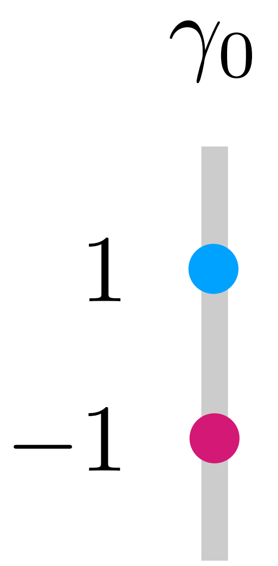
$$y_t = \theta^T x_t$$

INPUTS

(Independent Variables)

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

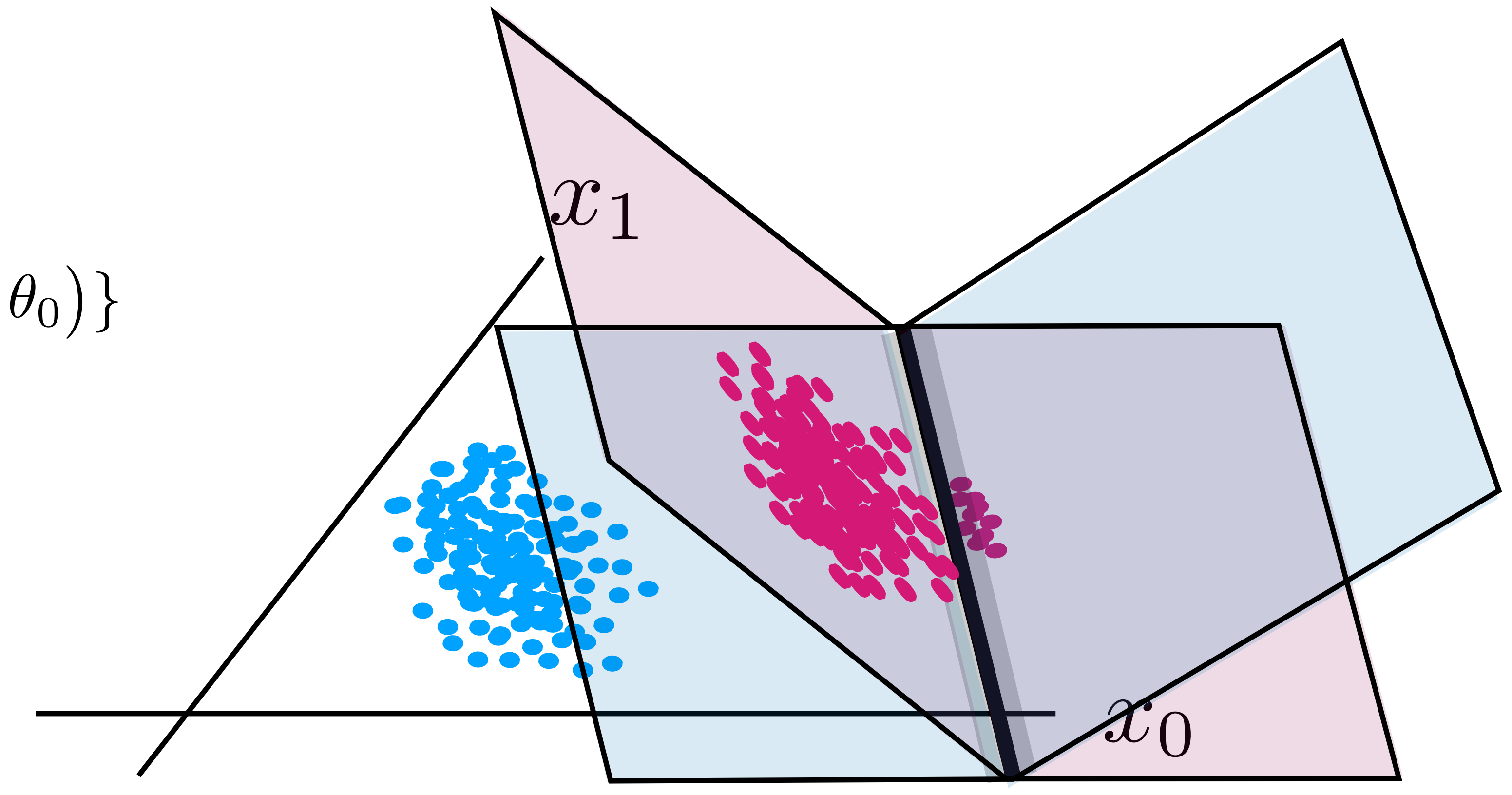
$$\gamma_t = \mathrm{sgn}(\theta^T x_t - \theta_0)$$



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

Support Vector Machine (SVM)

Soft boundary



“Binary classifier”

Classification

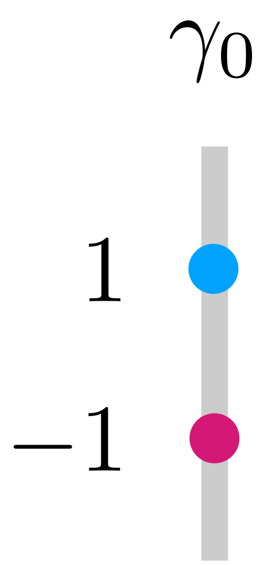
OUTPUTS
(Dependent Variables)

$$y_t = \theta^T x_t$$

INPUTS
(Independent Variables)

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

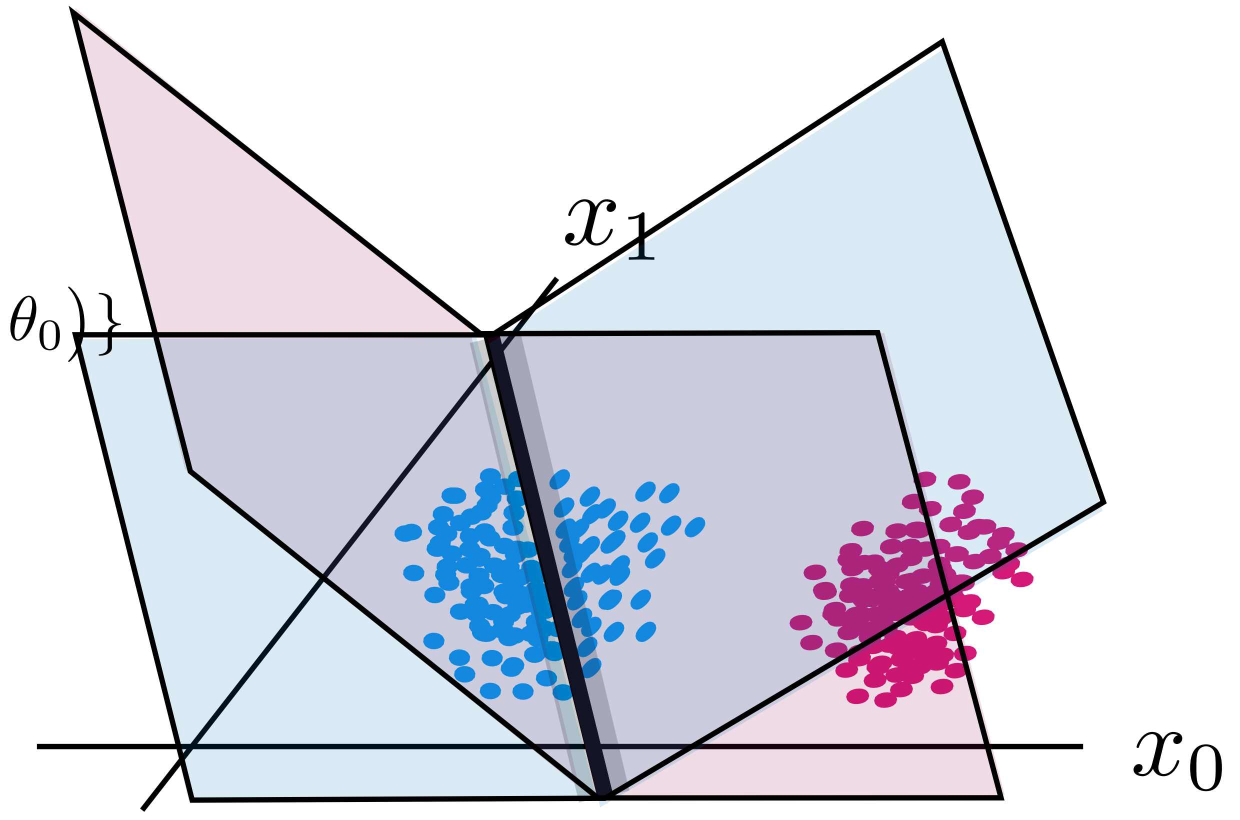
$$\gamma_t = \mathrm{sgn}(\theta^T x_t - \theta_0)$$



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

Support Vector Machine (SVM)

Soft boundary



“Binary classifier”

Classification

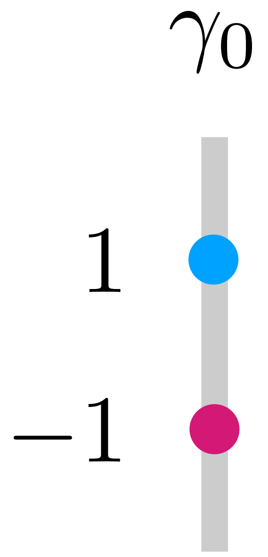
OUTPUTS
(Dependent Variables)

$$y_t = \theta^T x_t$$

INPUTS
(Independent Variables)

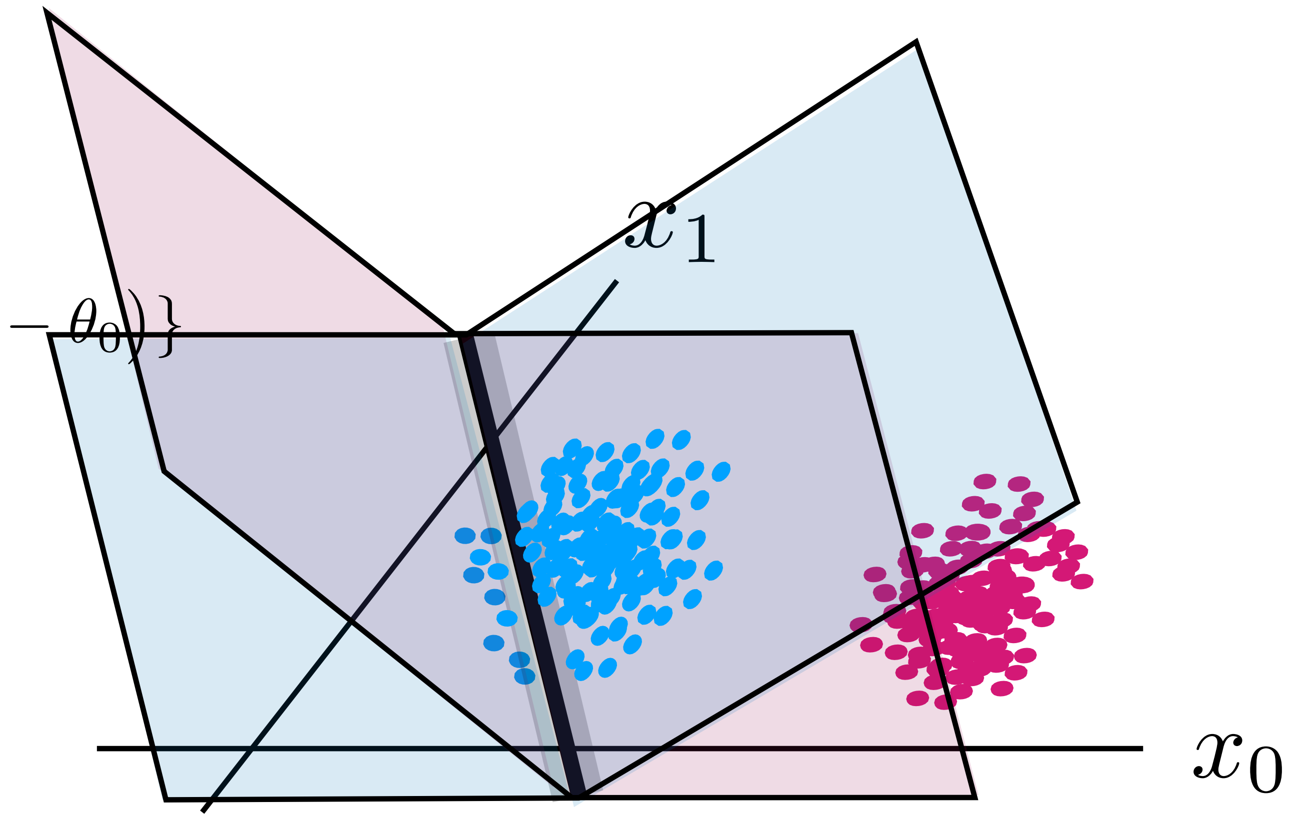
$$\begin{bmatrix} y_{00} & \cdots & y_{0m} & \gamma_{00} & \cdots & \gamma_{0m'} \\ y_{10} & \cdots & y_{1m} & \gamma_{10} & \cdots & \gamma_{1m'} \\ y_{20} & \cdots & y_{2m} & \gamma_{20} & \cdots & \gamma_{2m'} \\ y_{30} & \cdots & y_{3m} & \gamma_{30} & \cdots & \gamma_{3m'} \\ y_{40} & \cdots & y_{4m}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

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



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

Support Vector Machine (SVM)



Soft boundary

“Binary classifier”

Classification

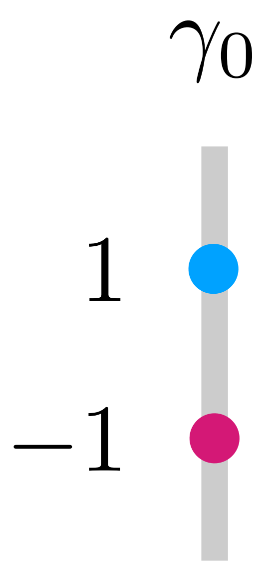
OUTPUTS
(Dependent Variables)

$$y_t = \theta^T x_t$$

INPUTS
(Independent Variables)

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

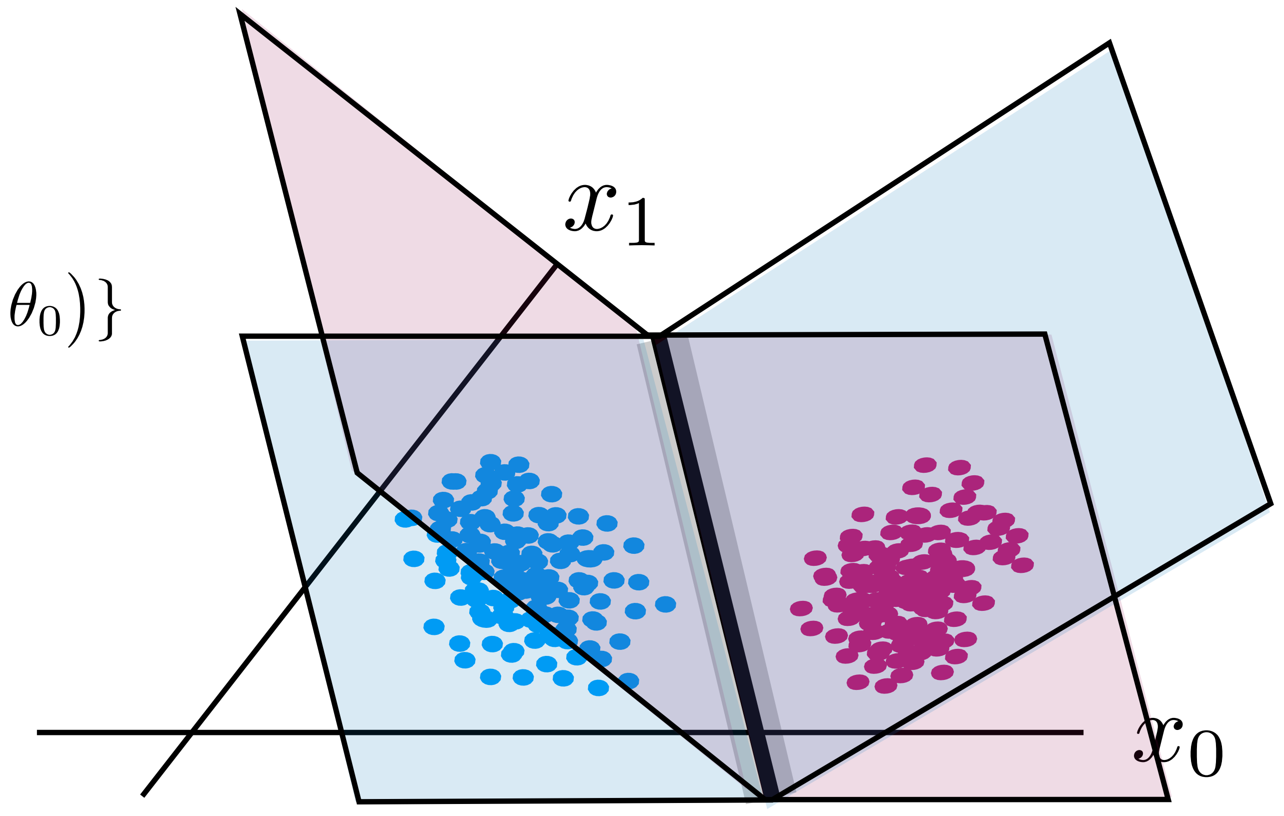
$$\gamma_t = \mathrm{sgn}(\theta^T x_t - \theta_0)$$



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

Support Vector Machine (SVM)

Soft boundary



"Binary classifier"

Classification

OUTPUTS

(Dependent Variables)

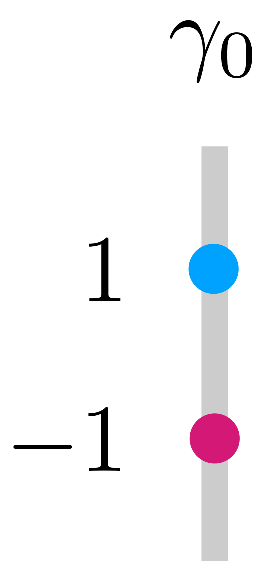
$$y_t = \theta^T x_t$$

INPUTS

(Independent Variables)

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

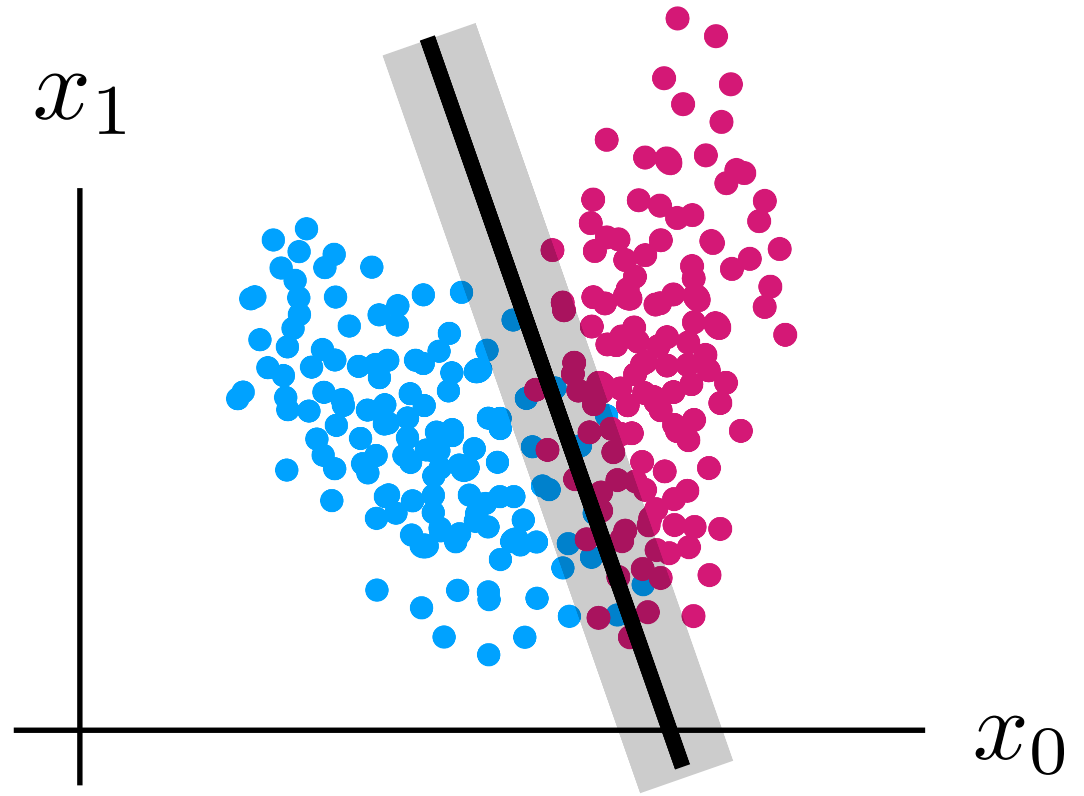
$$\gamma_t = \mathrm{sgn}(\theta^T x_t - \theta_0)$$



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

Support Vector Machine (SVM)

Soft boundary



“Binary classifier”

Classification

OUTPUTS

(Dependent Variables)

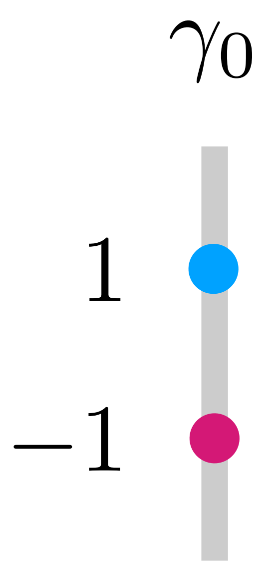
$$y_t = \theta^T x_t$$

INPUTS

(Independent Variables)

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

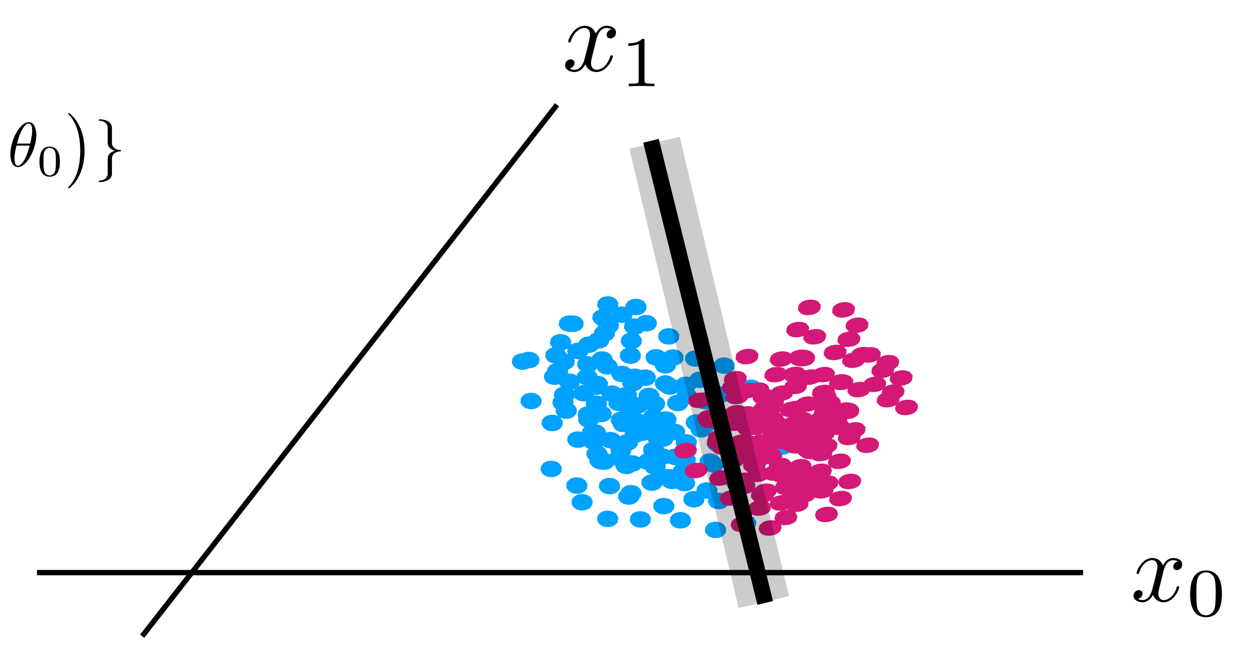
$$\gamma_t = \mathrm{sgn}(\theta^T x_t - \theta_0)$$



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

Support Vector Machine (SVM)

Soft boundary



“Binary classifier”

Classification

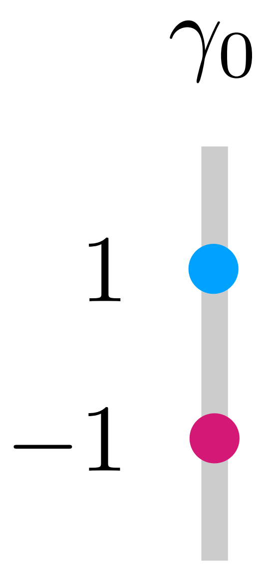
OUTPUTS
(Dependent Variables)

$$y_t = \theta^T x_t$$

INPUTS
(Independent Variables)

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

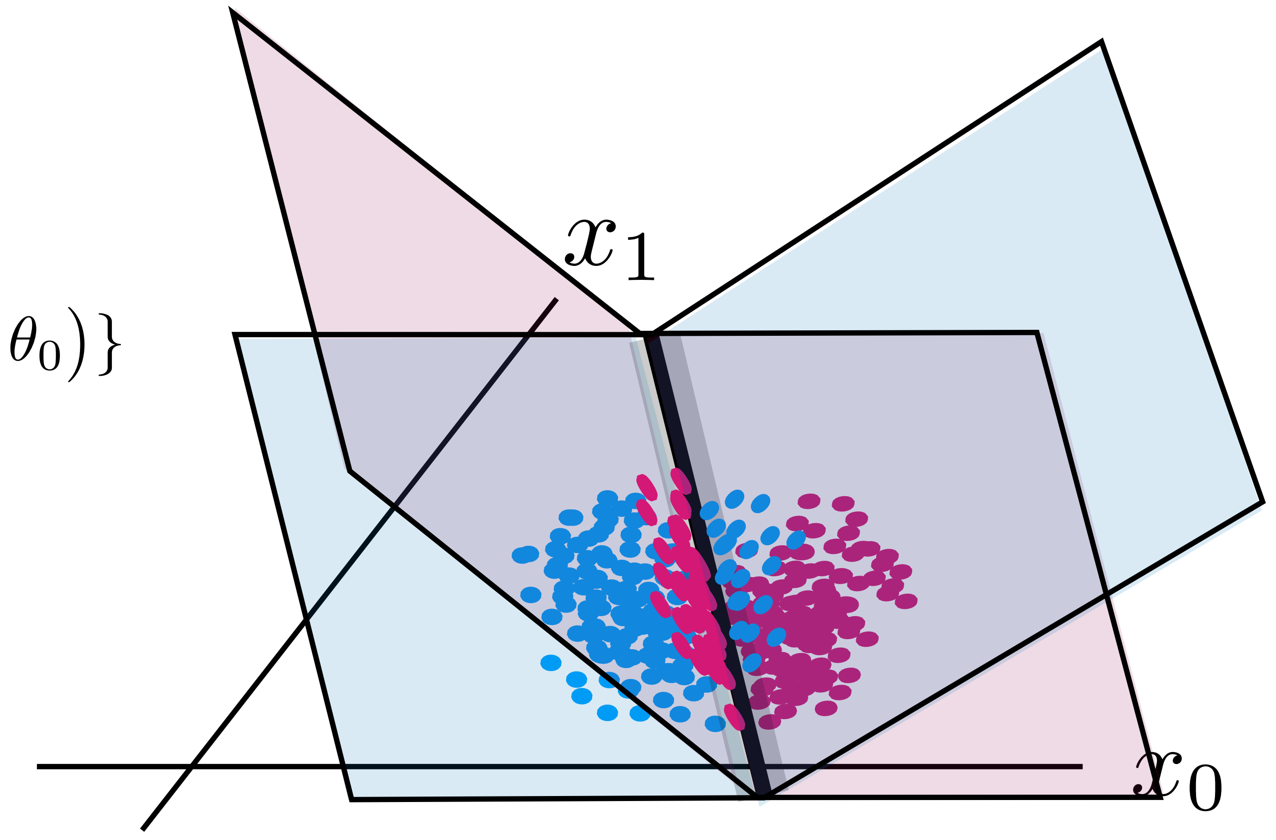
$$\gamma_t = \mathrm{sgn}(\theta^T x_t - \theta_0)$$



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

Support Vector Machine (SVM)

Soft boundary



“Binary classifier”

Classification

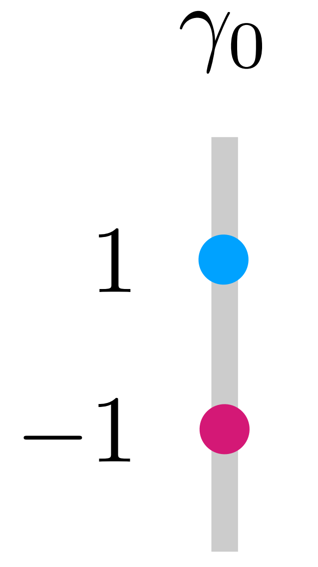
OUTPUTS
(Dependent Variables)

$$y_t = \theta^T x_t$$

INPUTS
(Independent Variables)

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

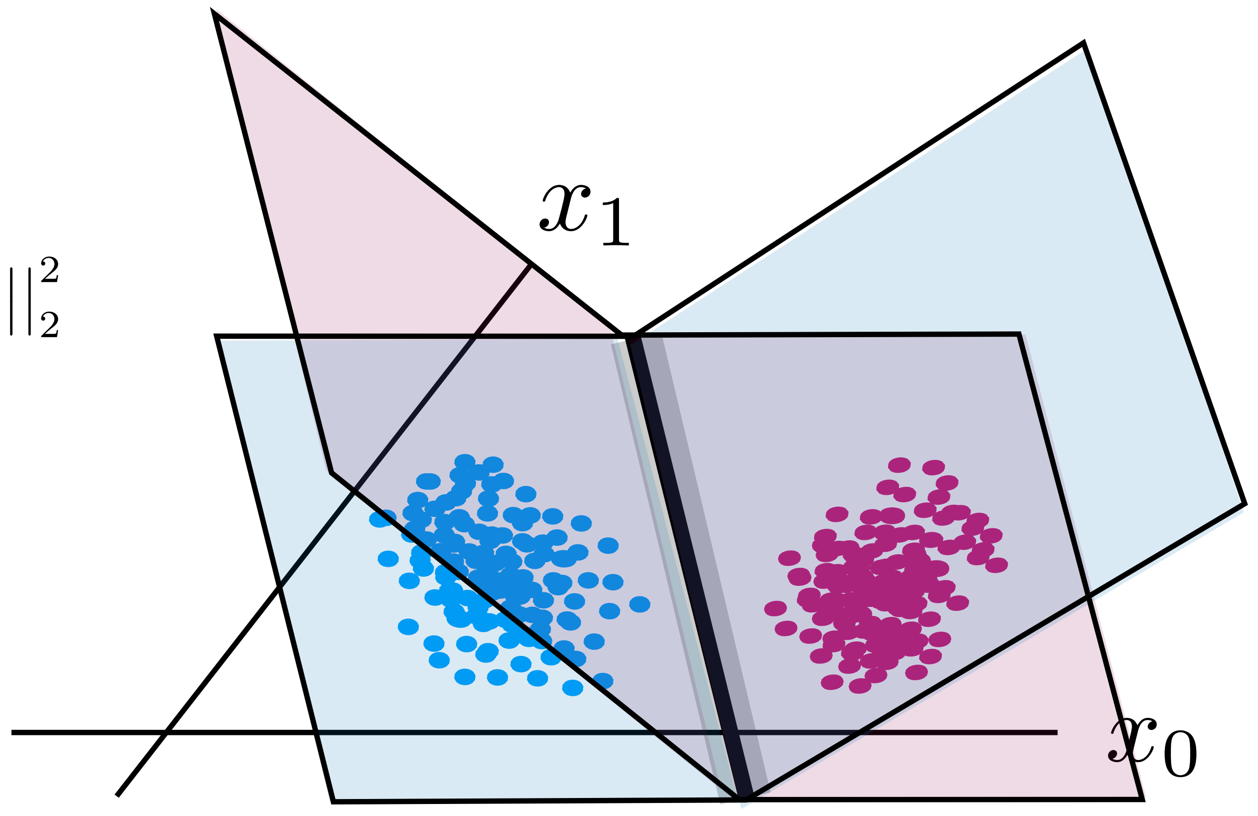
$$\gamma_t = \operatorname{sgn}(\theta^T x_t - \theta_0)$$



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

Support Vector Machine (SVM)

Soft boundary



"Binary classifier"

Classification

OUTPUTS

(Dependent Variables)

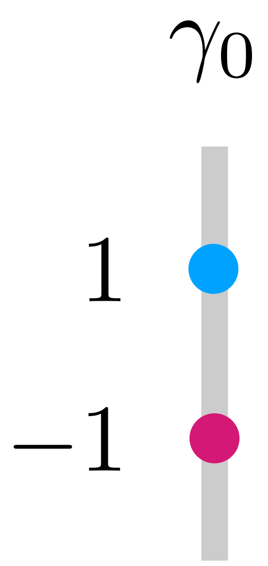
$$y_t = \theta^T x_t$$

INPUTS

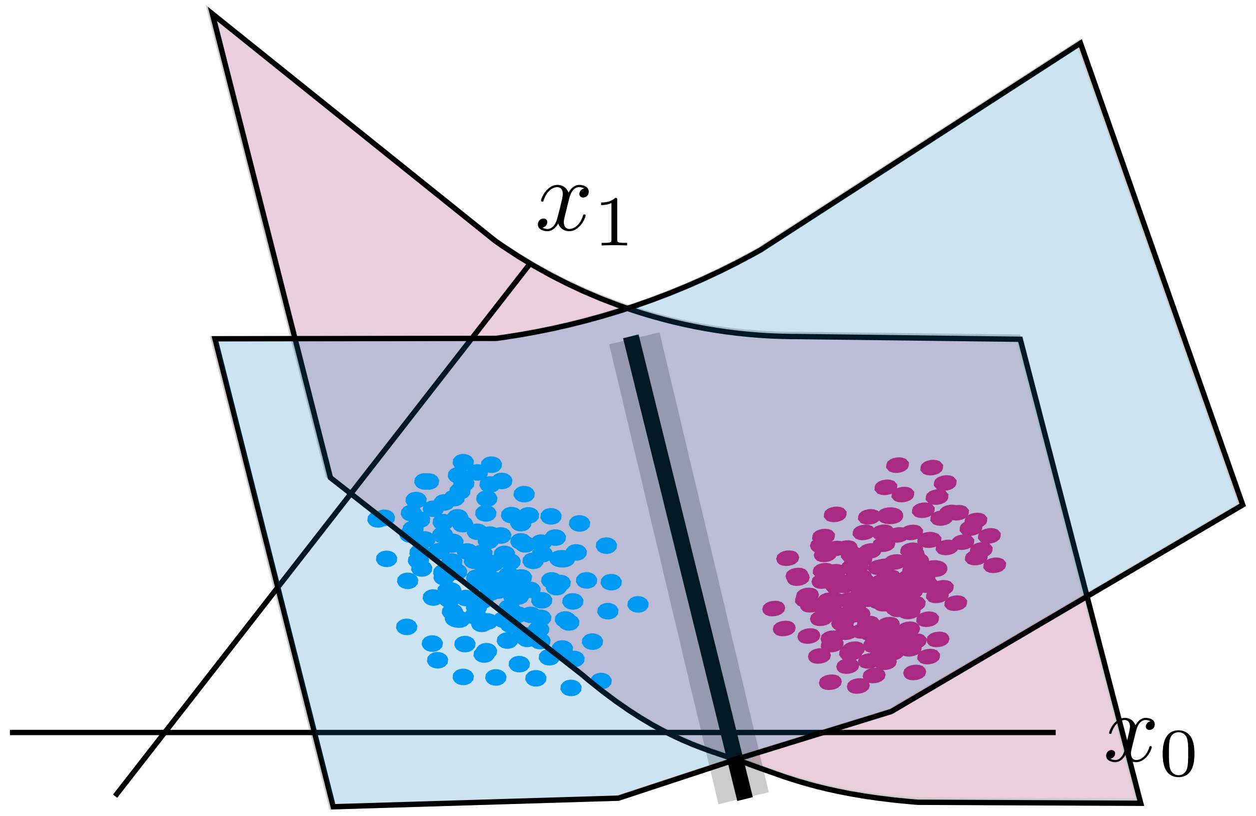
(Independent Variables)

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

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



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



Support Vector Machine (SVM)

Logistic Regression (smoothed)

Soft boundary

“Binary classifier”

Classification

OUTPUTS

(Dependent Variables)

$$y_t = \theta^T x_t$$

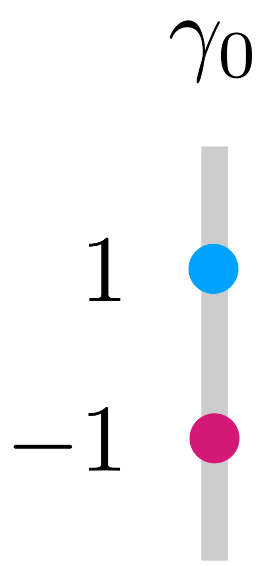
INPUTS

(Independent Variables)

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

with basis functions

$$\gamma_t = \operatorname{sgn}\left(\theta^T h_t(x_t) - \theta_0\right)$$



COST:

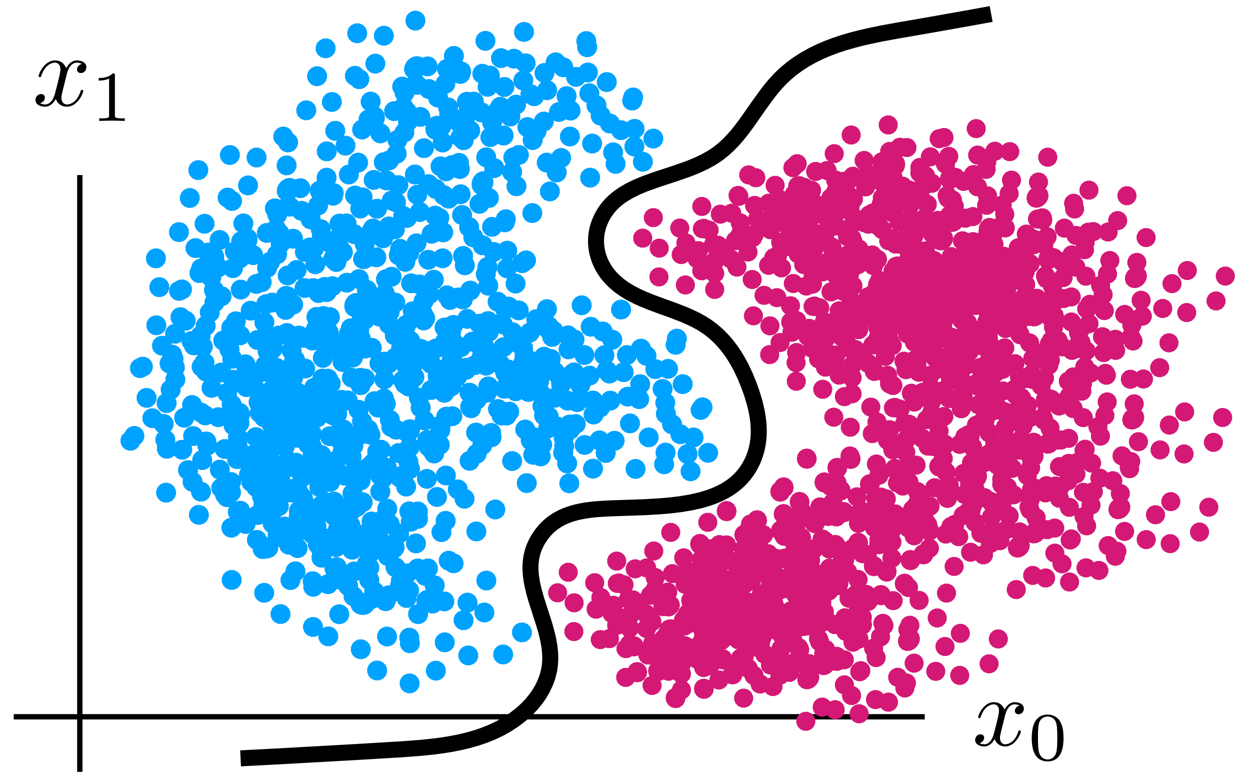
$$\min_\theta \quad \frac{1}{T}\sum_t \max\left\{0, 1 - \gamma_t\left(\theta_{1:n}^T h_t(x_t) - \theta_0\right)\right\} + \lambda||\theta_{1:n}||_2^2$$

Support Vector Machine (SVM)

Logistic Regression (smoothed)

Soft boundary

"Binary classifier"



Classification

OUTPUTS
(Dependent Variables)

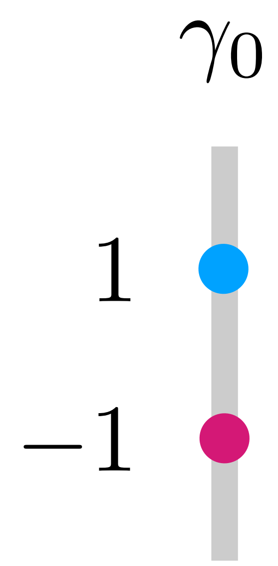
$$y_t = \theta^T x_t$$

INPUTS
(Independent Variables)

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

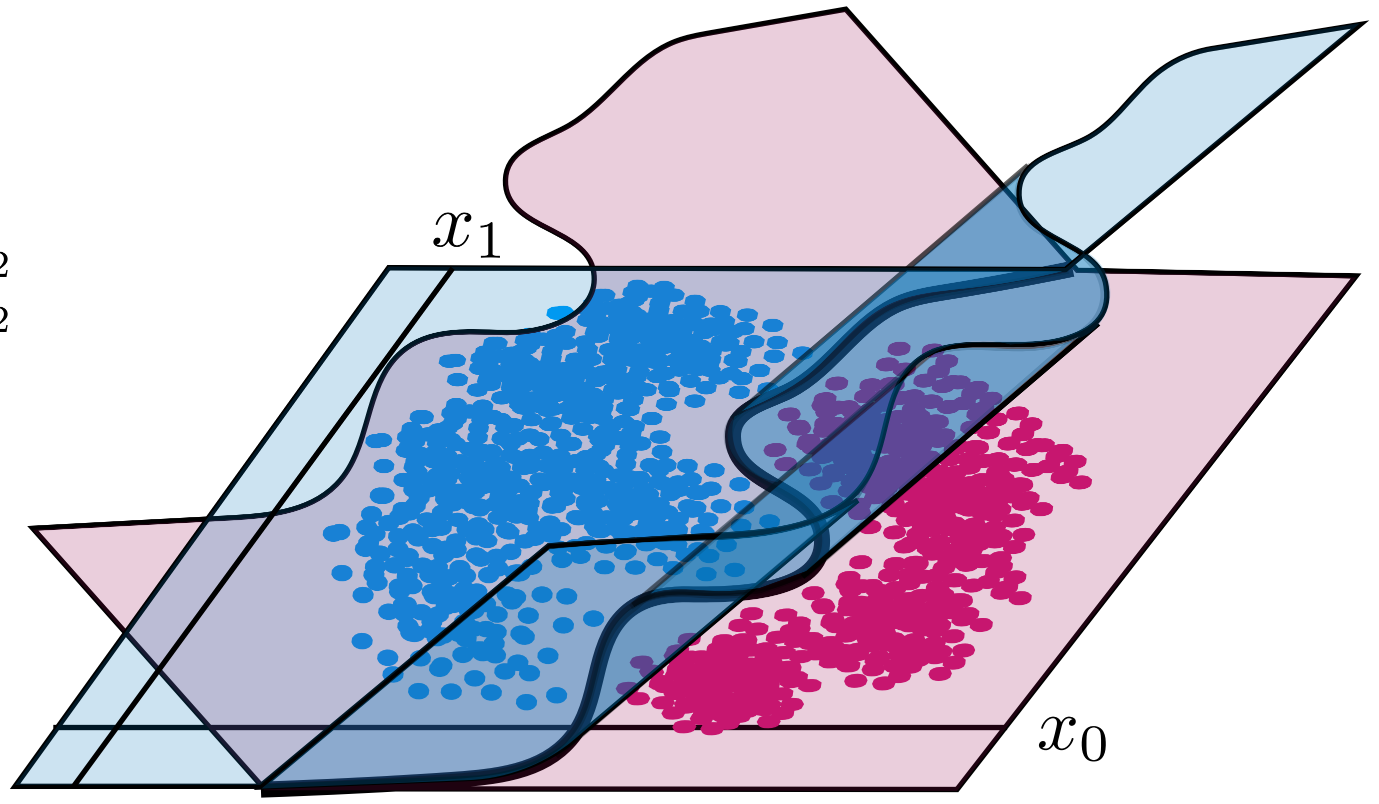
with basis functions

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



COST:

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



Support Vector Machine (SVM)

Hard boundary

Logistic Regression (smoothed)

"Binary classifier"

Classification

OUTPUTS

(Dependent Variables)

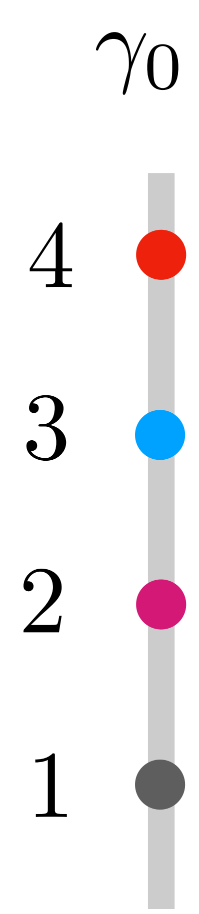
$$y_t = \theta^T x_t$$

INPUTS

(Independent Variables)

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

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



COST:

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

Multi-class

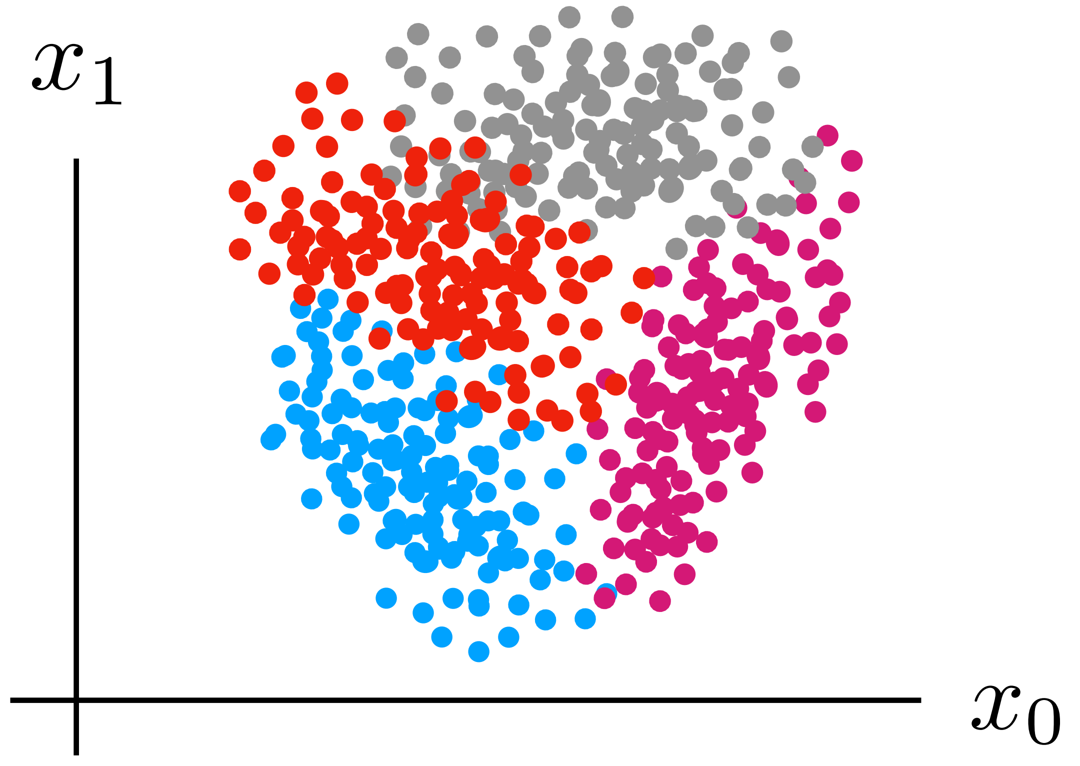
Support Vector Machine (SVM)

Logistic Regression (smoothed)

Collection of Binary Classifiers

“One to the others”

# of classifiers: m



Classification

OUTPUTS

(Dependent Variables)

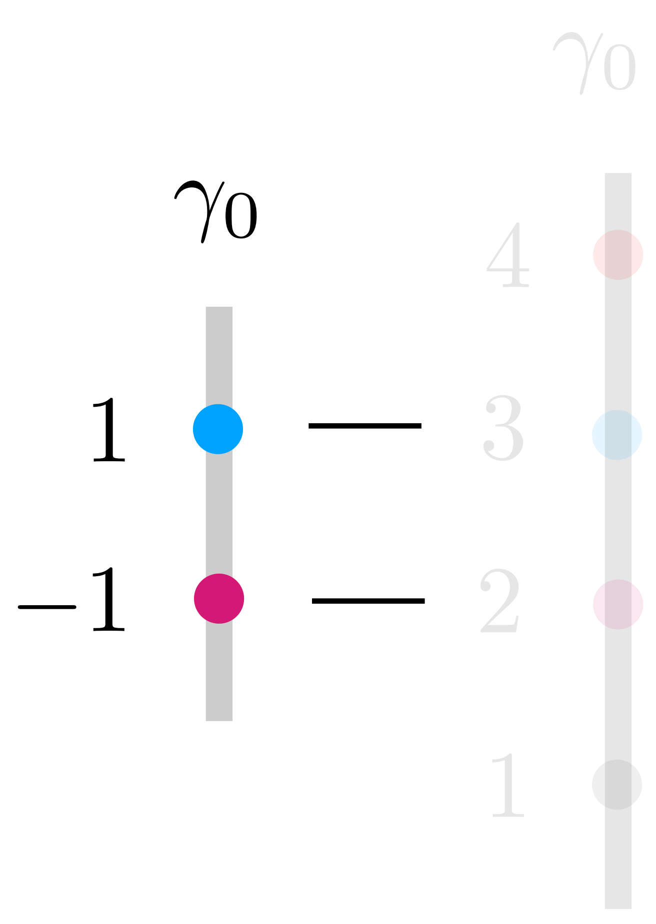
$$y_t = \theta^T x_t$$

INPUTS

(Independent Variables)

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

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



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

Multi-class

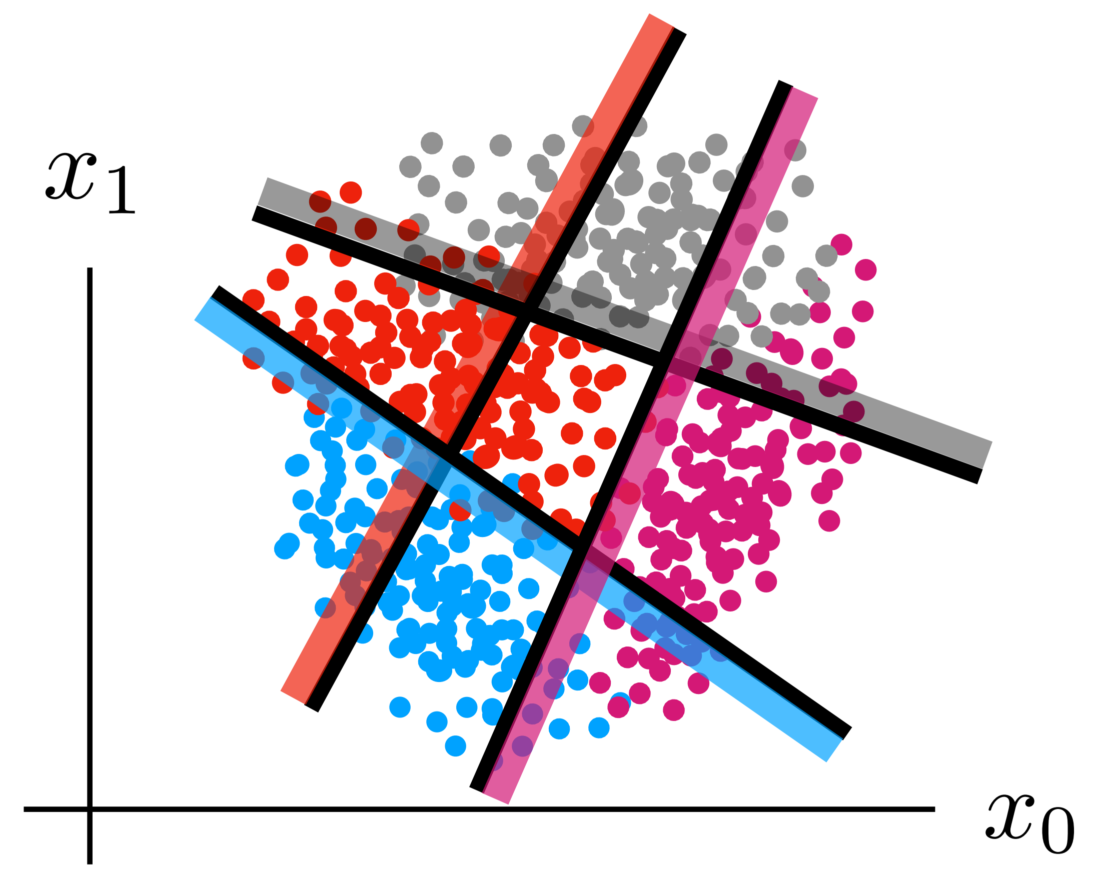
Support Vector Machine (SVM)

Logistic Regression (smoothed)

Collection of Binary Classifiers

“One to the others”

# of classifiers: m



Classification

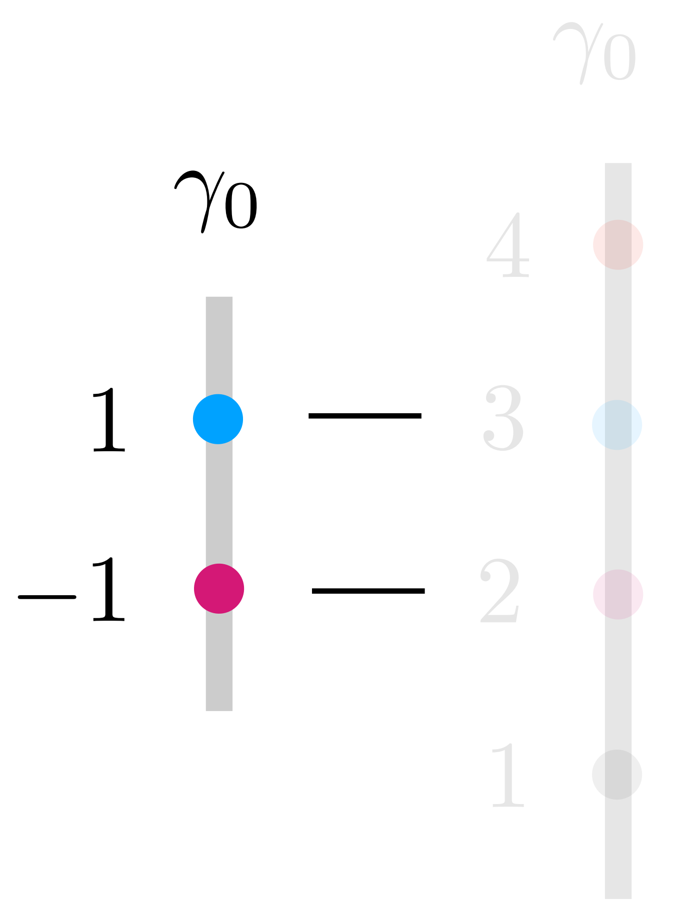
OUTPUTS
(Dependent Variables)

$$y_t = \theta^T x_t$$

INPUTS
(Independent Variables)

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

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



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

Multi-class

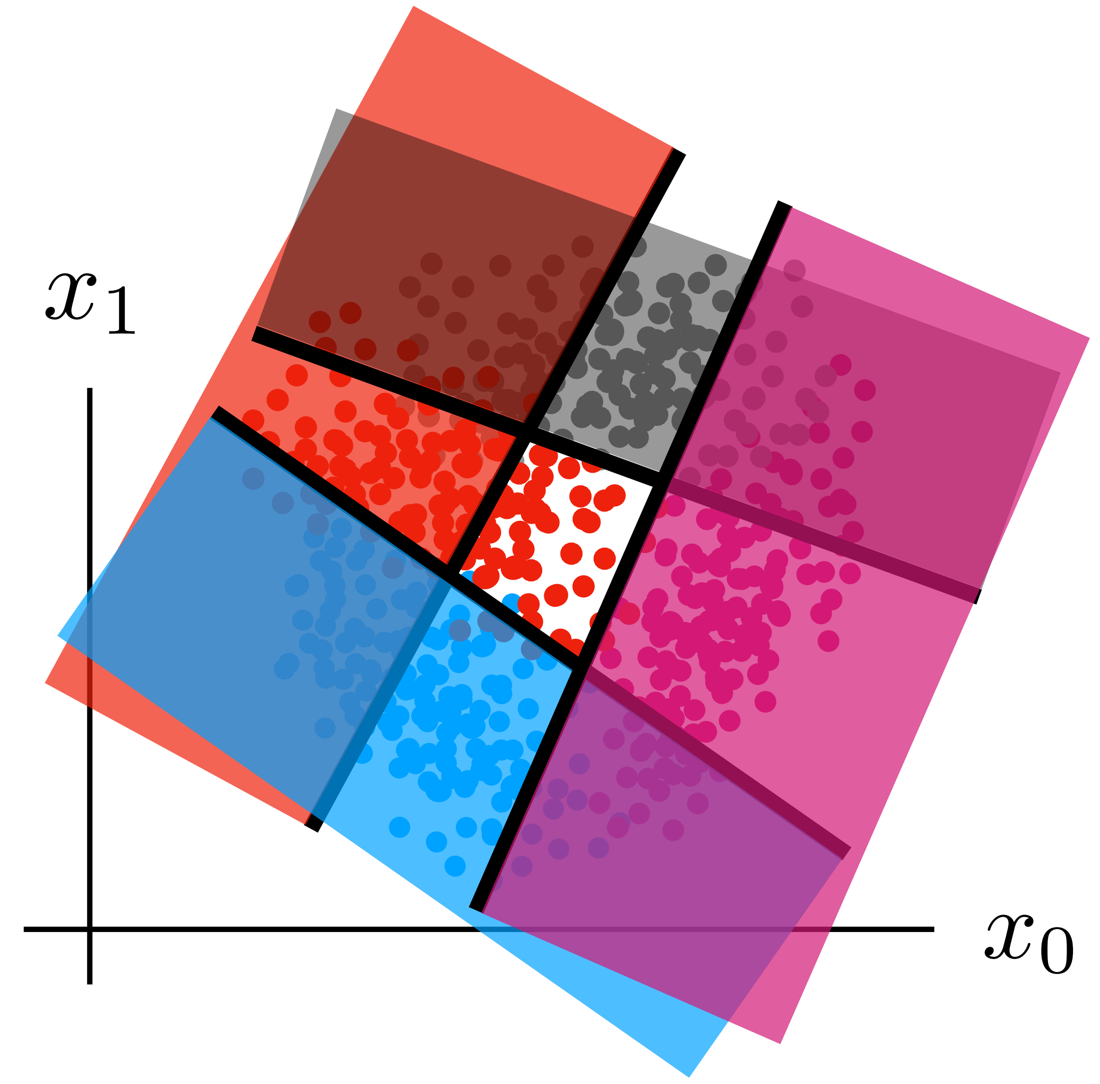
Support Vector Machine (SVM)

Logistic Regression (smoothed)

Collection of Binary Classifiers

“One to the others”

# of classifiers: m



Classification

OUTPUTS

(Dependent Variables)

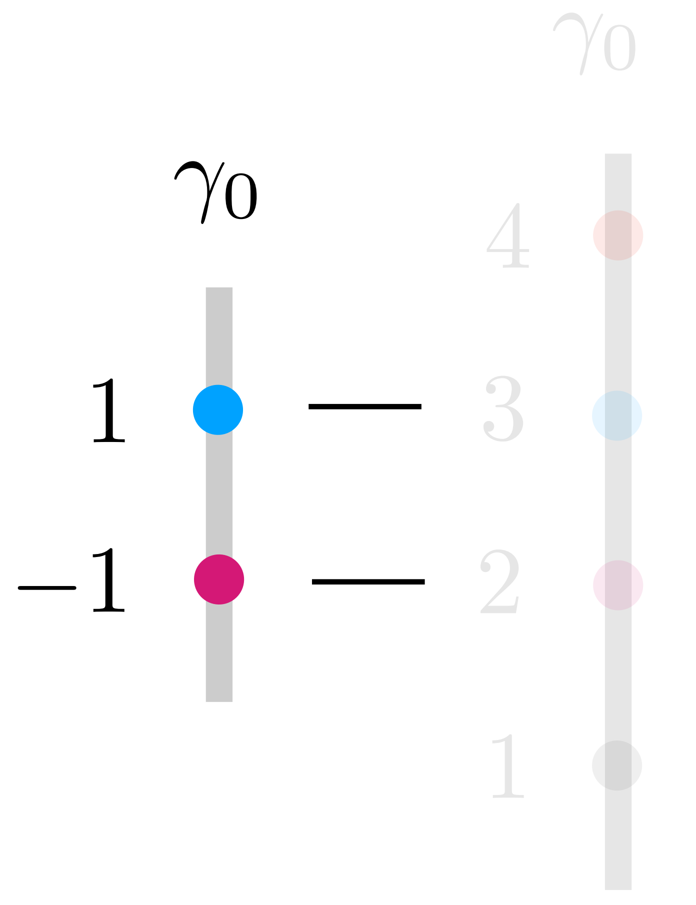
$$y_t = \theta^T x_t$$

INPUTS

(Independent Variables)

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

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



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

Multi-class

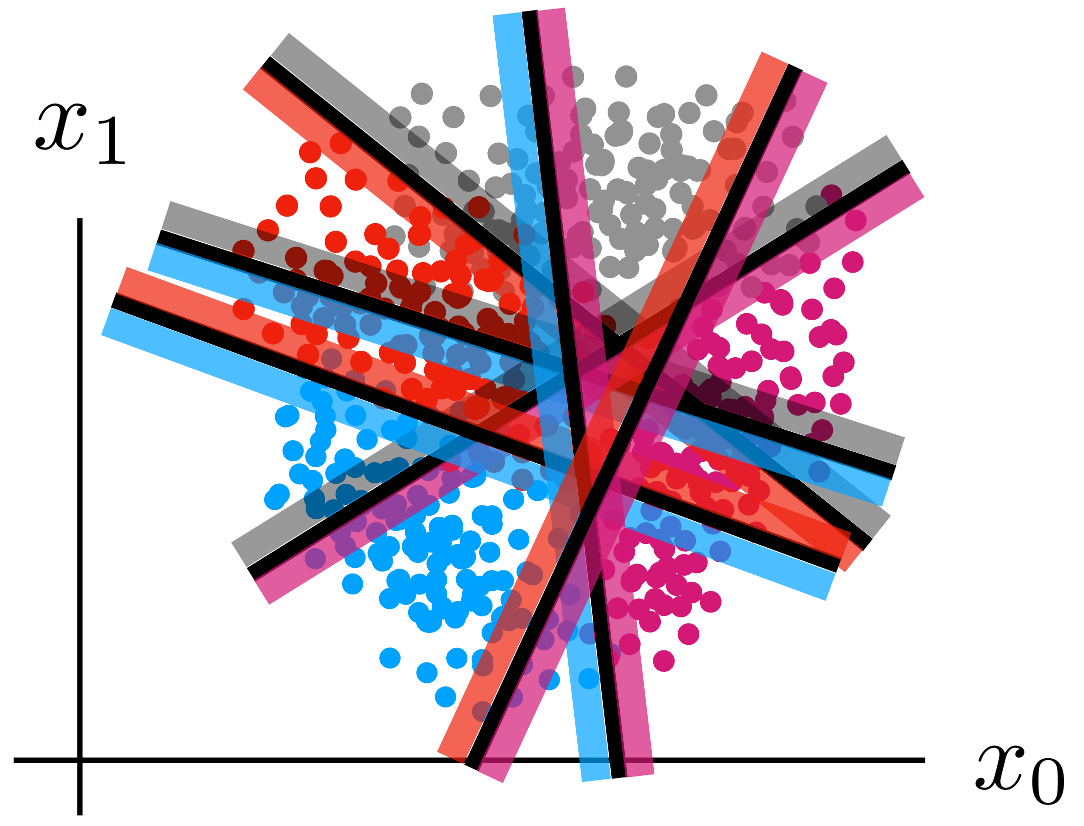
Support Vector Machine (SVM)

Logistic Regression (smoothed)

Collection of Binary Classifiers

Pairwise - “One to another”

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



Classification

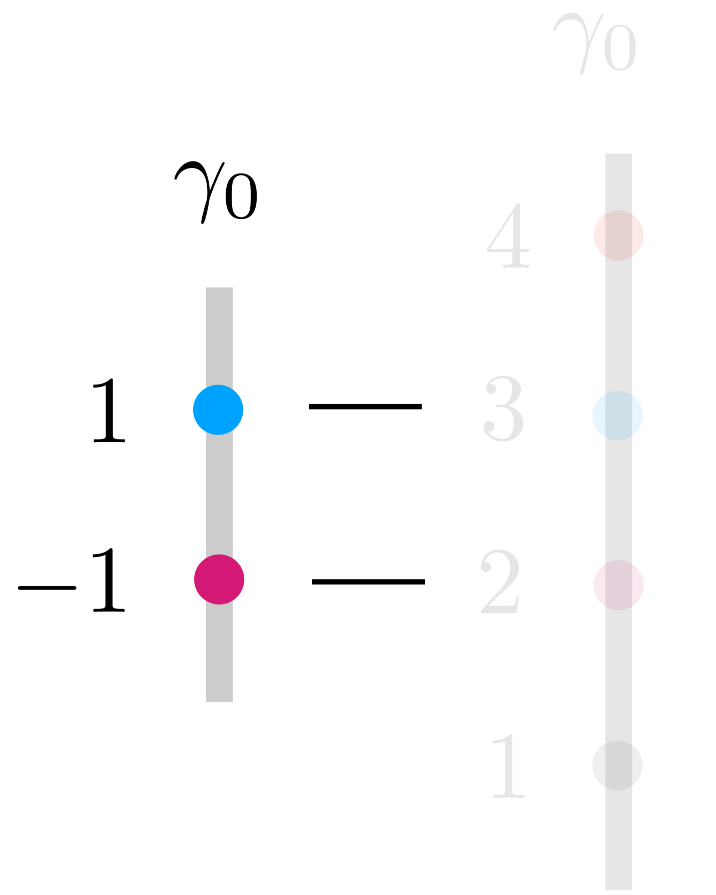
OUTPUTS
(Dependent Variables)

$$y_t = \theta^T x_t$$

INPUTS
(Independent Variables)

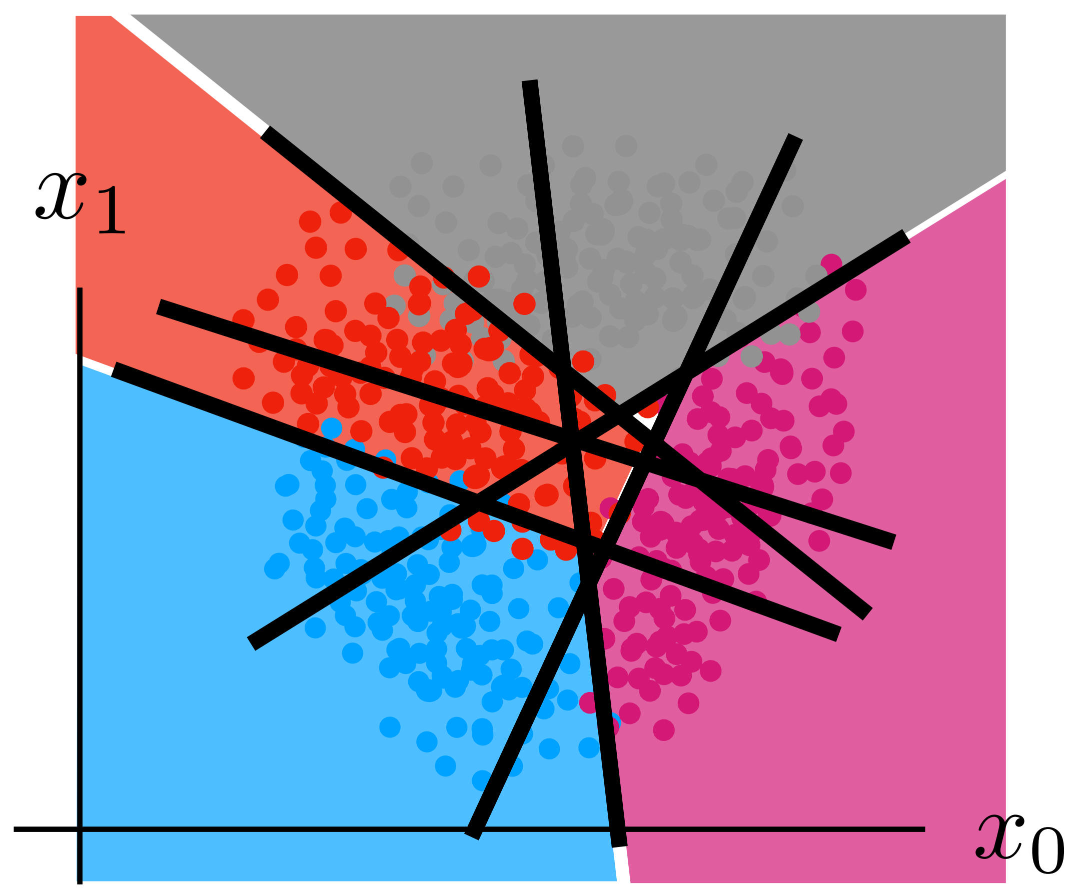
$$\begin{bmatrix} y_{00} & \cdots & y_{0m} & \gamma_{00} & \cdots & \gamma_{0m'} \\ y_{10} & \cdots & y_{1m} & \gamma_{10} & \cdots & \gamma_{1m'} \\ y_{20} & \cdots & y_{2m} & \gamma_{20} & \cdots & \gamma_{2m'} \\ y_{30} & \cdots & y_{3m} & \gamma_{30} & \cdots & \gamma_{3m'} \\ y_{40} & \cdots & y_{4m}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

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



COST:

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



Multi-class

Support Vector Machine (SVM)

Logistic Regression (smoothed)

Collection of Binary Classifiers

Pairwise - “One to another”

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

Classification

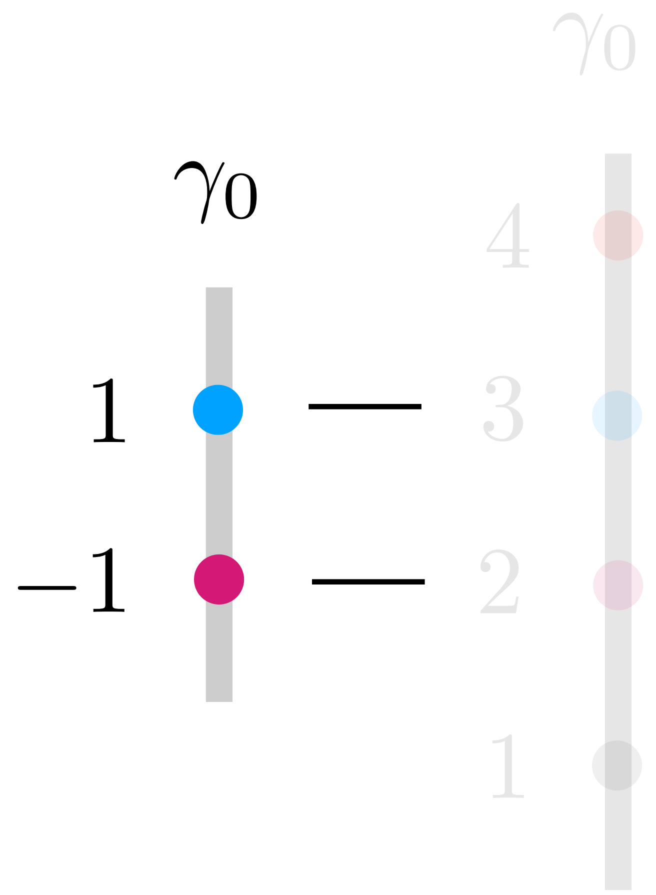
OUTPUTS
(Dependent Variables)

$$y_t = \theta^T x_t$$

INPUTS
(Independent Variables)

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

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



COST:

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

Multi-class

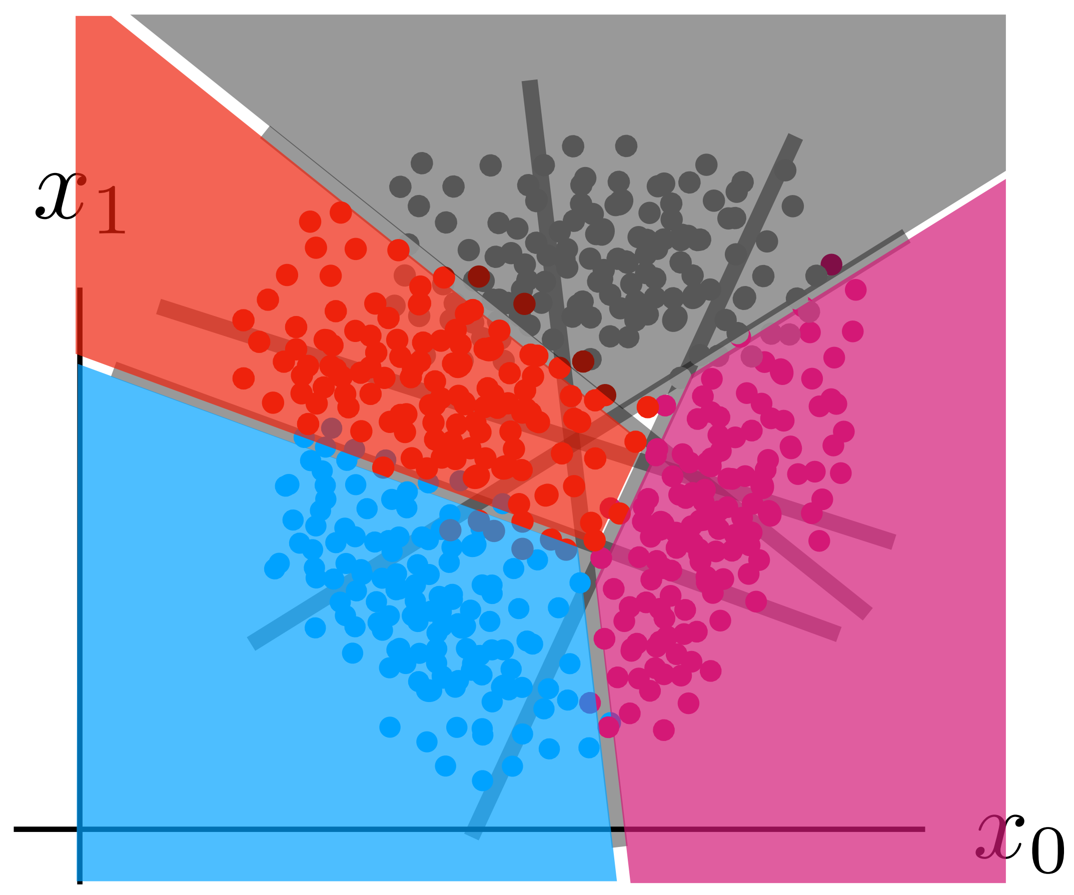
Support Vector Machine (SVM)

Logistic Regression (smoothed)

Collection of Binary Classifiers

Pairwise - “One to another”

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



Classification

OUTPUTS
(Dependent Variables)

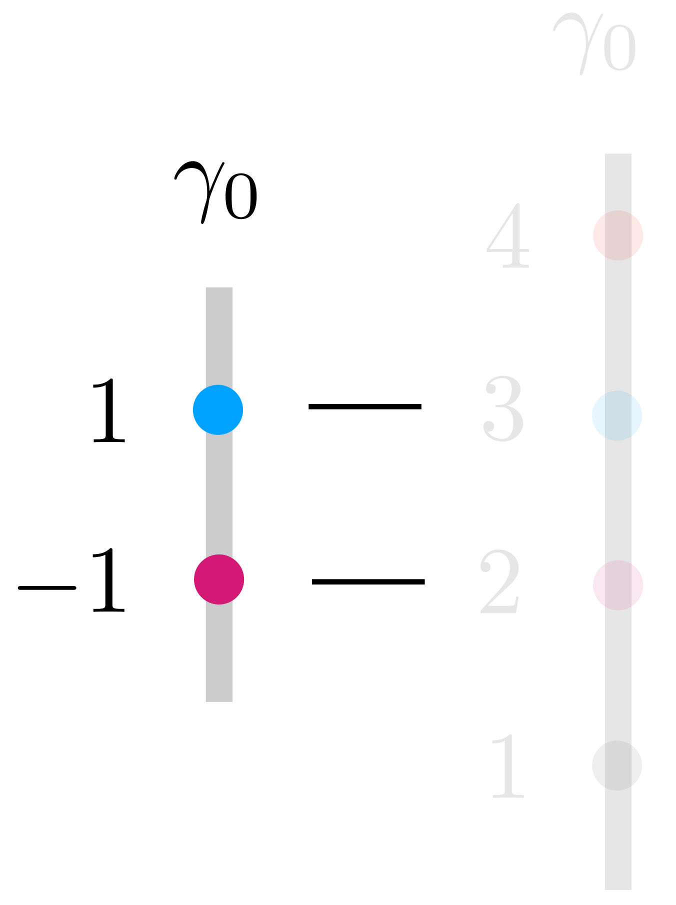
$$y_t = \theta^T x_t$$

INPUTS
(Independent Variables)

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

with BASIS FUNCTIONS

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



COST:

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

Multi-class

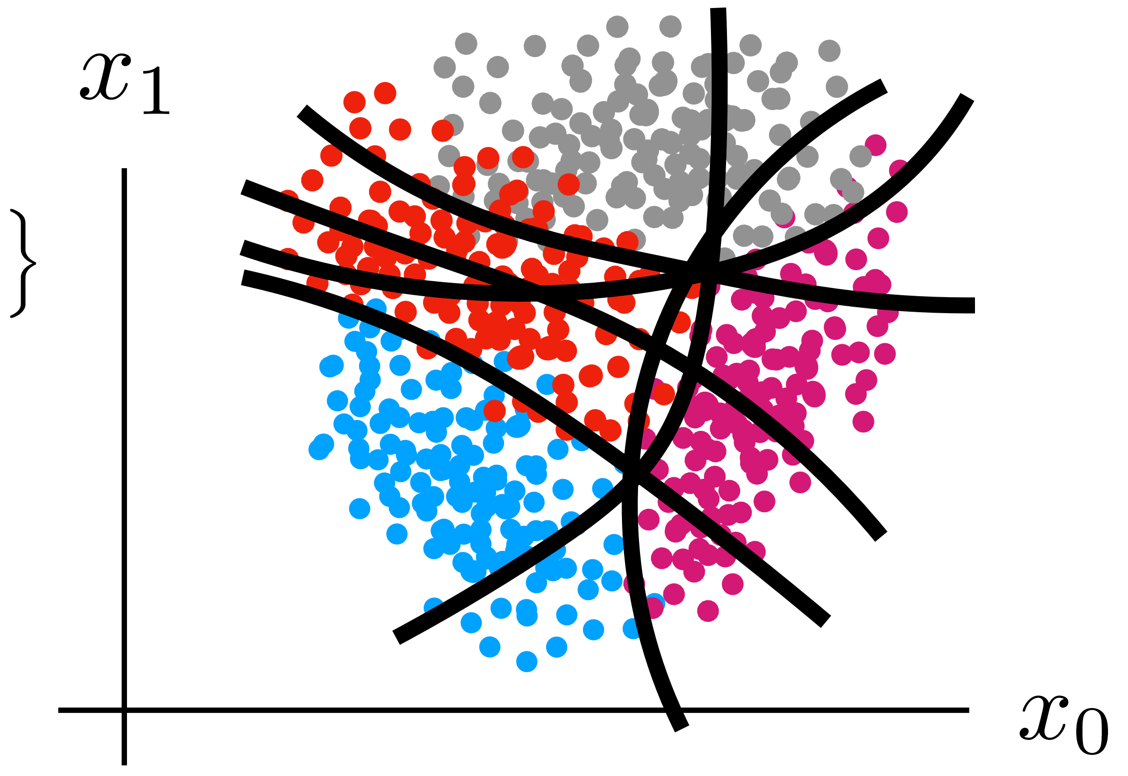
Support Vector Machine (SVM)

Logistic Regression (smoothed)

Collection of Binary Classifiers

Pairwise - "One to another"

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



Classification

OUTPUTS
(Dependent Variables)

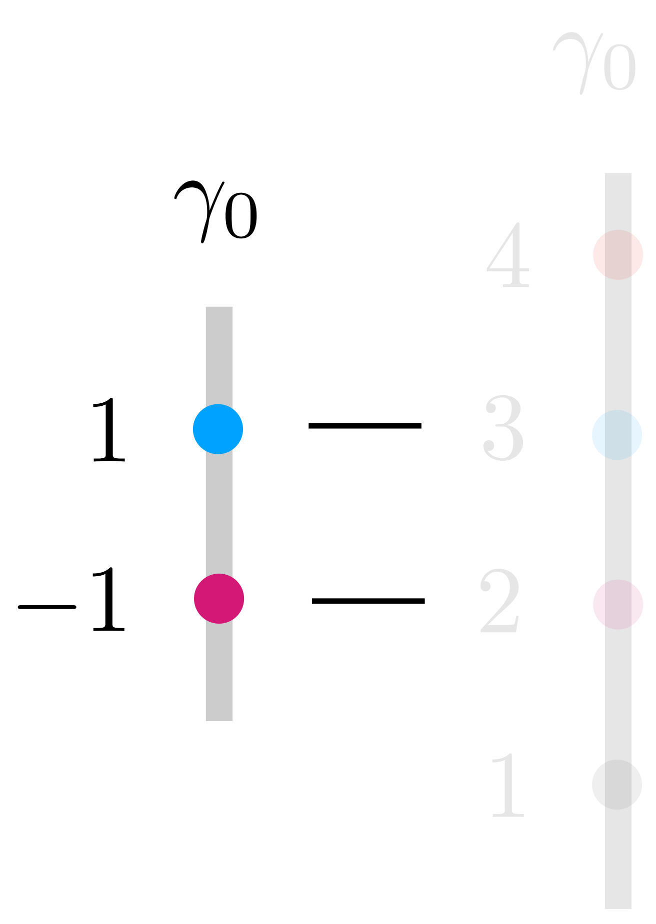
$$y_t = \theta^T x_t$$

INPUTS
(Independent Variables)

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

with BASIS FUNCTIONS

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



COST:

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

Multi-class

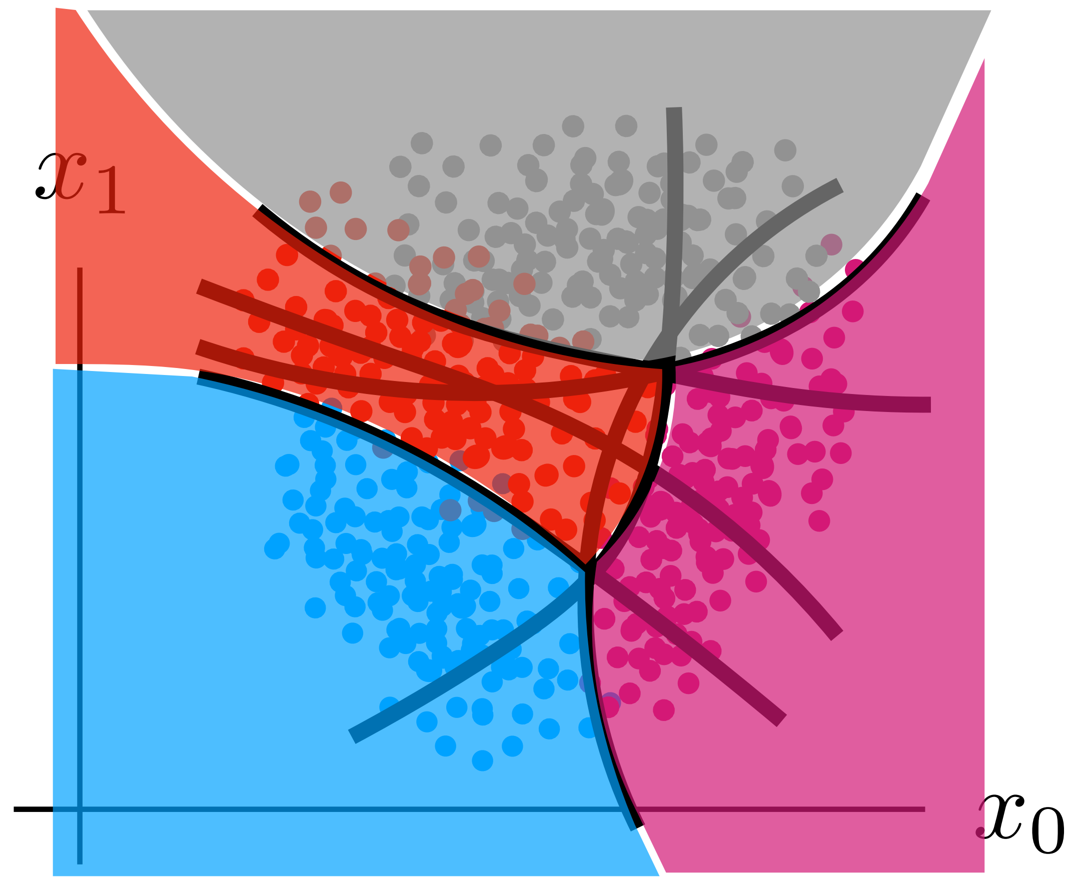
Support Vector Machine (SVM)

Logistic Regression (smoothed)

Collection of Binary Classifiers

Pairwise - “One to another”

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



Classification

OUTPUTS
(Dependent Variables)

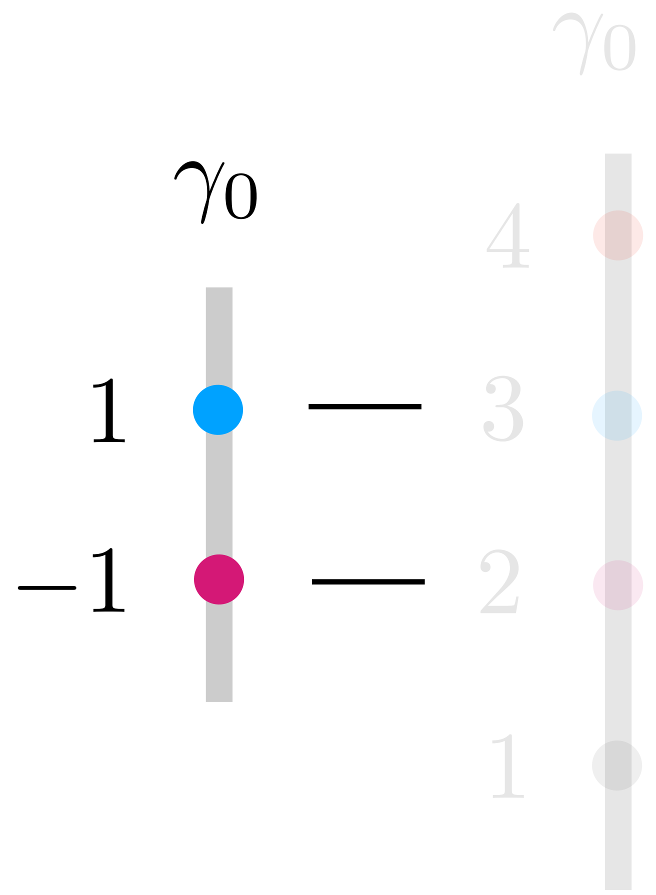
$$y_t = \theta^T x_t$$

INPUTS
(Independent Variables)

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

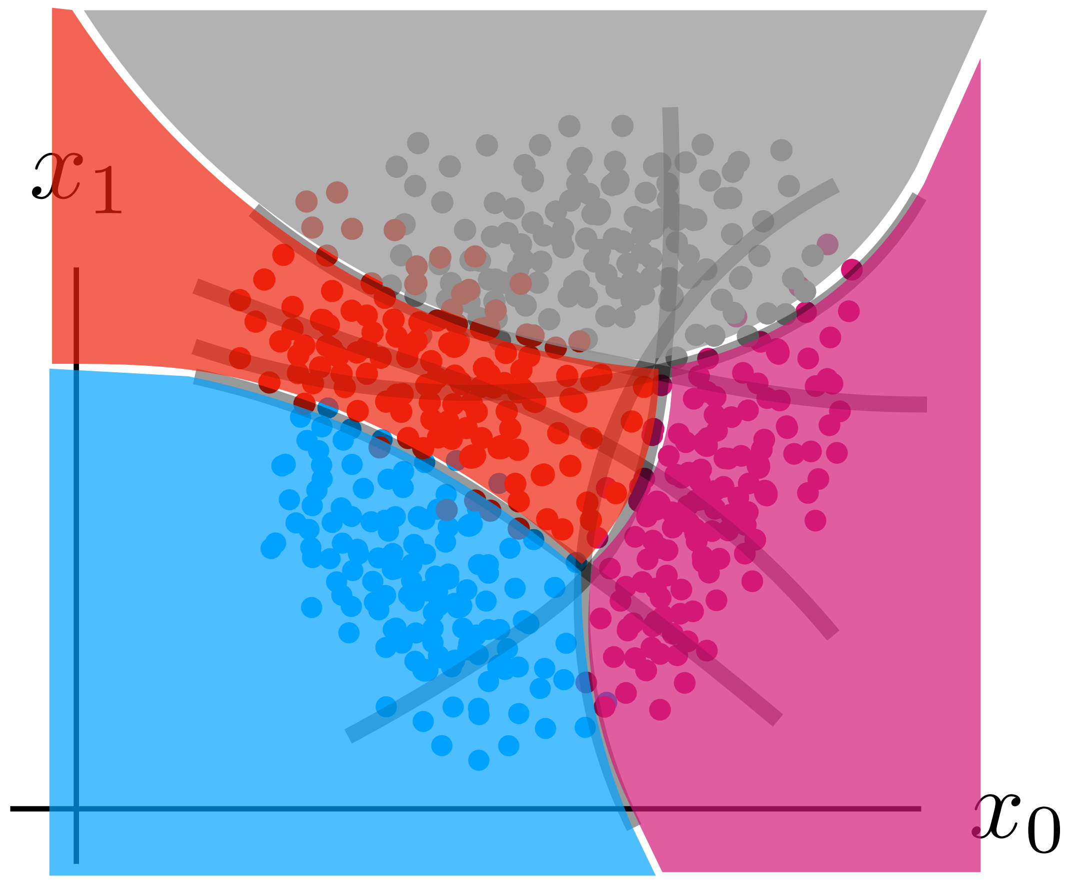
with BASIS FUNCTIONS

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



COST:

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



Multi-class

Support Vector Machine (SVM)

Logistic Regression (smoothed)

Collection of Binary Classifiers

Pairwise - "One to another"

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

Discriminant Analysis

OUTPUTS
(Dependent Variables)

$$y_t = \theta^T x_t$$

INPUTS
(Independent Variables)

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

CONDITIONAL PROBABILITY

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

BAYES RULE

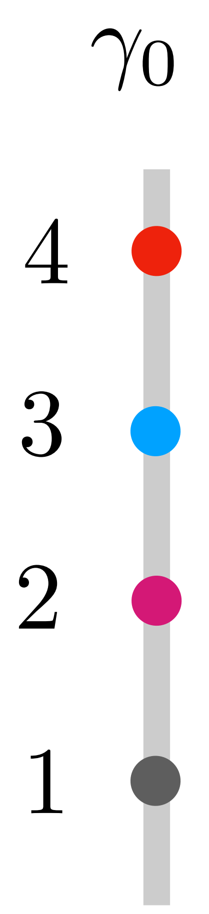
$$P(y|x) = \frac{P(x|y)P(y)}{P(x)}$$

$P(x|y=k)$ …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

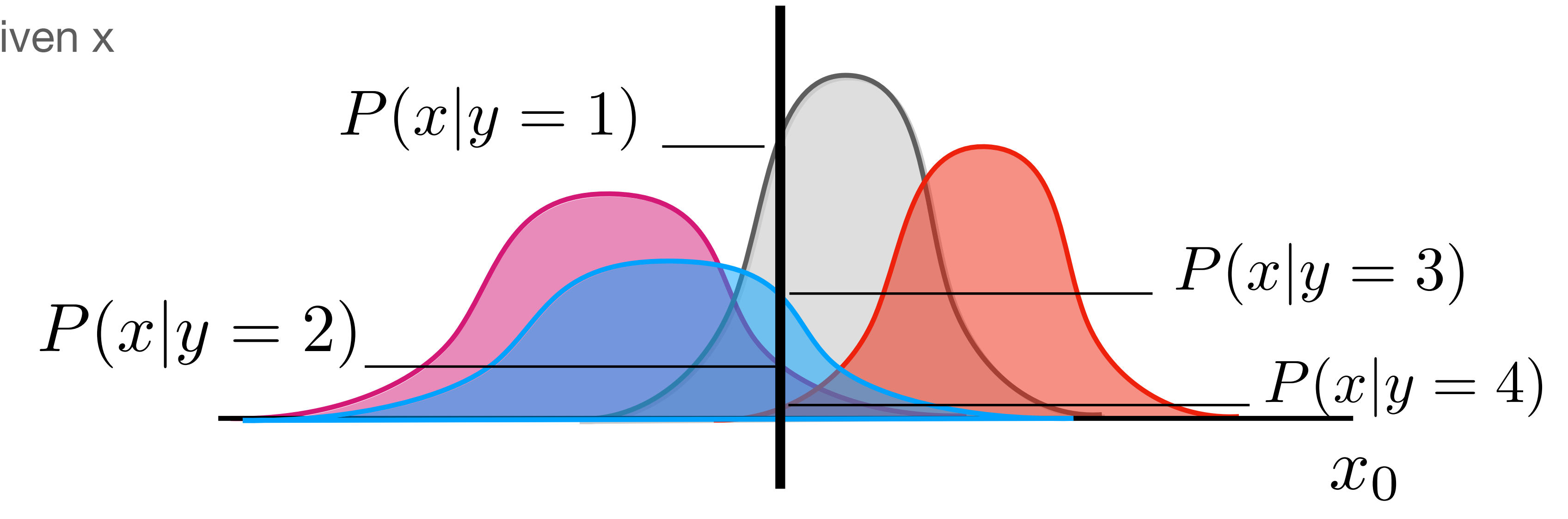


Class score

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

Predicted class

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



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

Discriminant Analysis

OUTPUTS
(Dependent Variables)

$$y_t = \theta^T x_t$$

INPUTS
(Independent Variables)

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

CONDITIONAL PROBABILITY

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

BAYES RULE

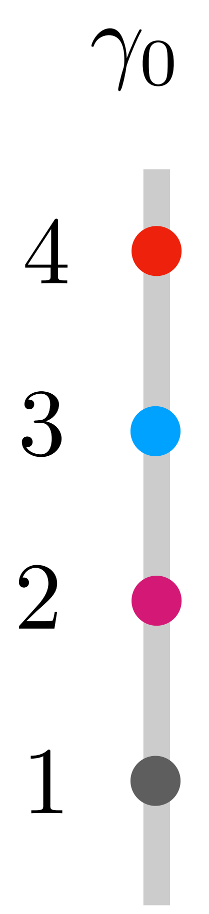
$$P(y|x) = \frac{P(x|y)P(y)}{P(x)}$$

$P(x|y=k)$ …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

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

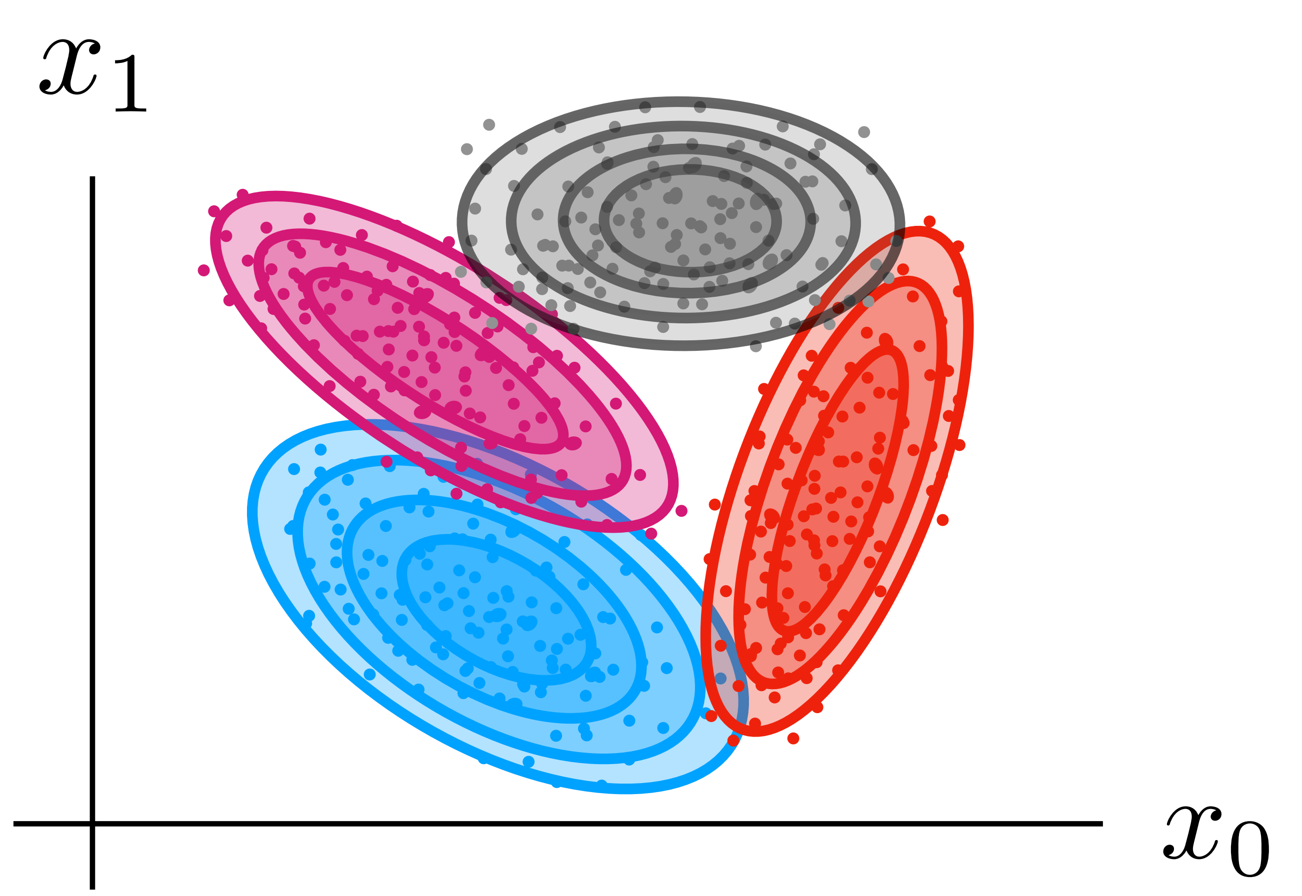
Discrimination Surfaces

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

Quadratic Discriminant Analysis

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

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)

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

CONDITIONAL PROBABILITY

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

BAYES RULE

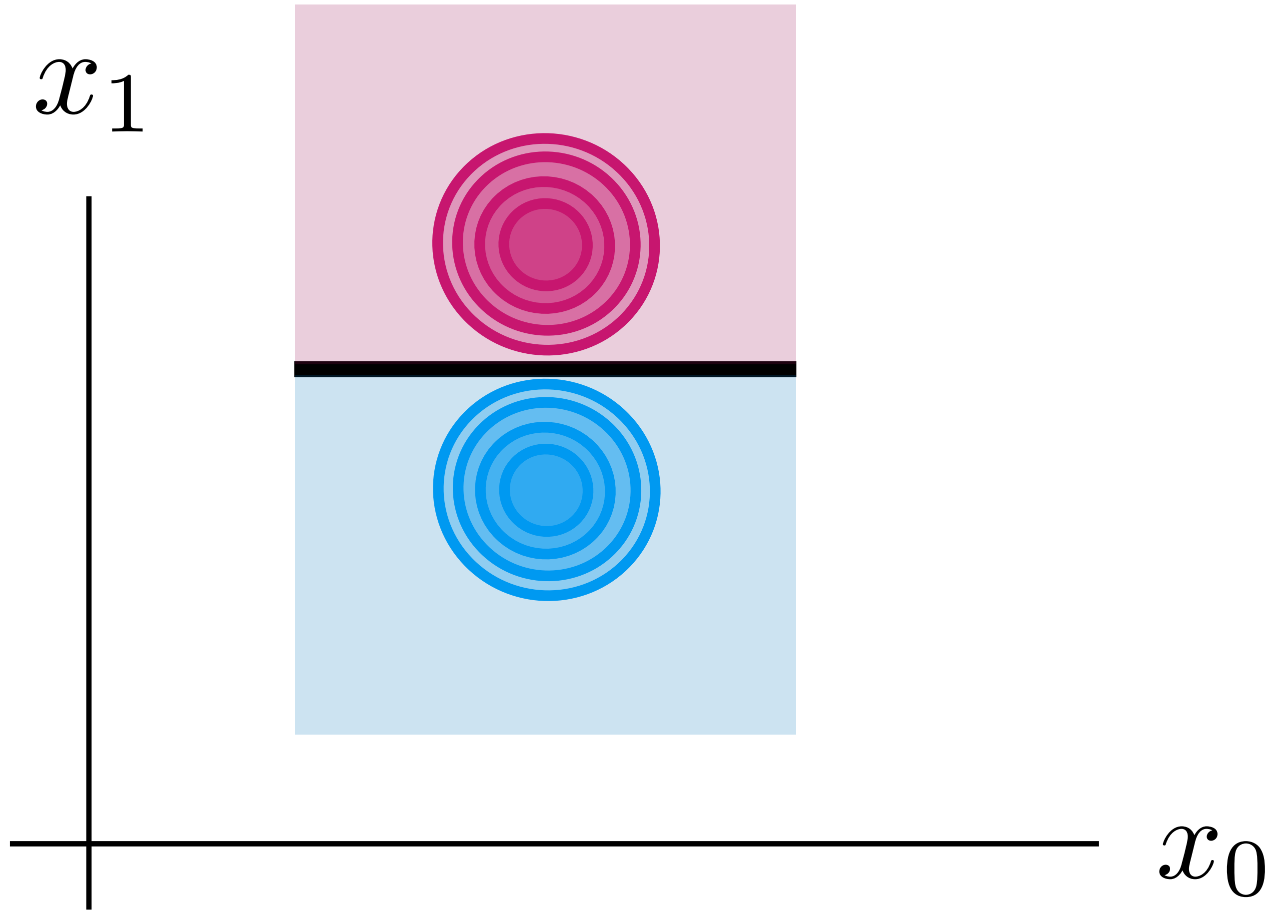
$$P(y|x) = \frac{P(x|y)P(y)}{P(x)}$$

$P(x|y=k)$ …height of density k

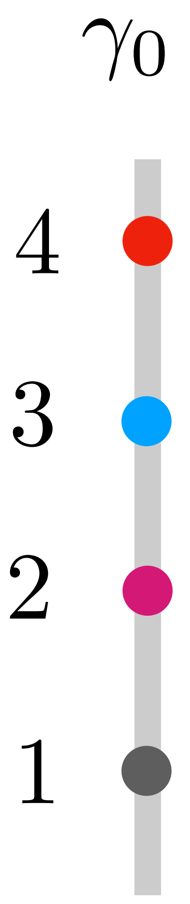
$P(y=k)$ …prior probability of k

$P(x)$ …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

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

ellipses or hyperbolas (or parabolas)

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)

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

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

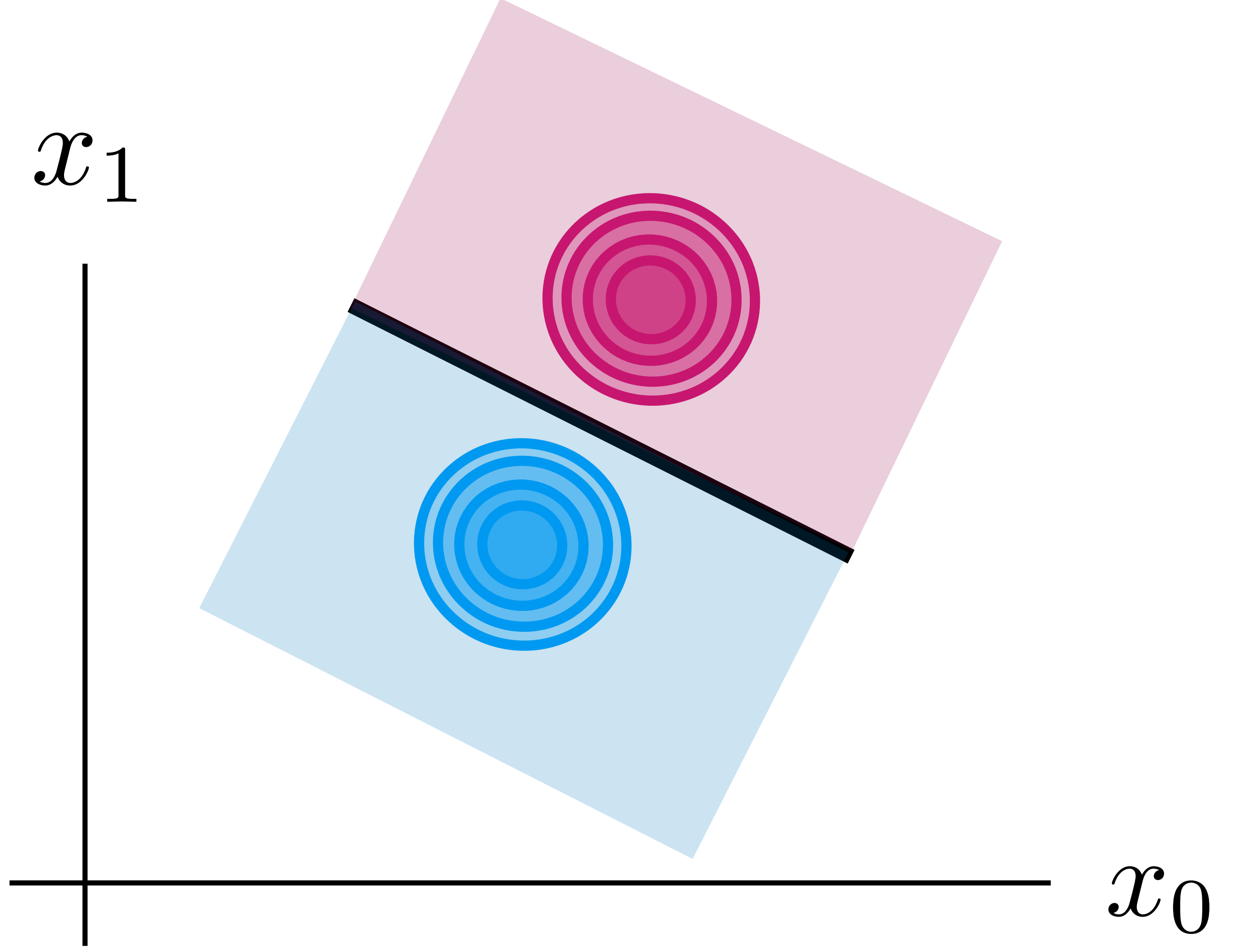
BAYES RULE $P(y|x) = \dfrac{P(x|y)P(y)}{P(x)}$

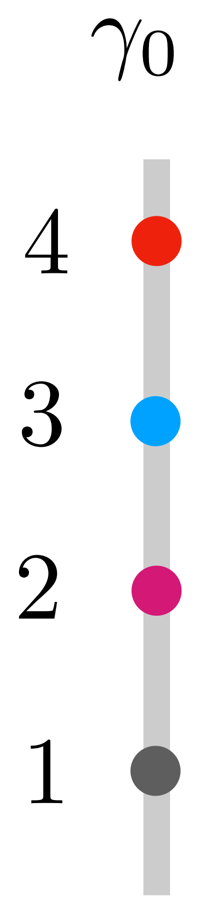
$P(x|y=k)$ …height of density k

$P(y=k)$ …prior probability of k

$P(x)$ …normalize factor

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





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

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

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

ellipses or hyperbolas (or parabolas)

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

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)

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

CONDITIONAL PROBABILITY

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

BAYES RULE

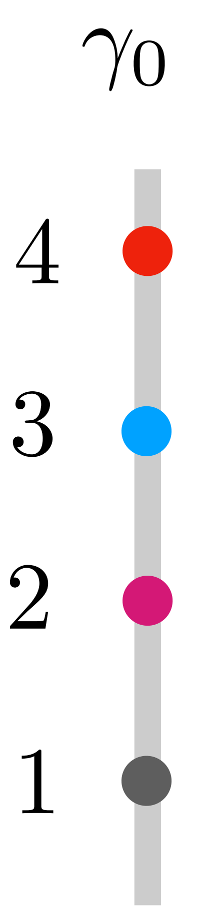
$$P(y|x) = \frac{P(x|y)P(y)}{P(x)}$$

$P(x|y=k)$ …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

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

Discrimination Surfaces

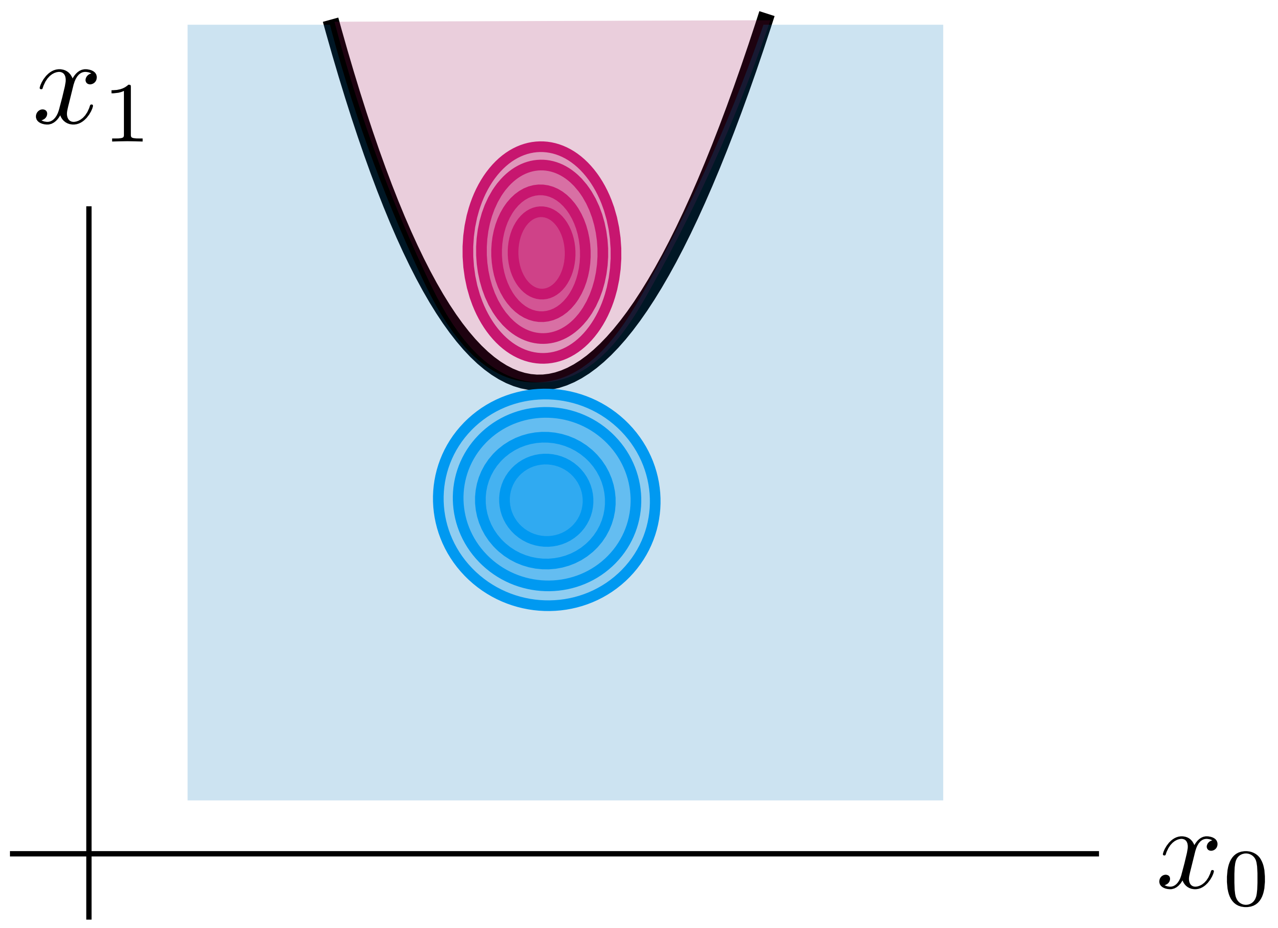
$$\log\ P(y=k|x) = \log\ P(y=k'|x)$$

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)

$$y_t = \theta^T x_t$$

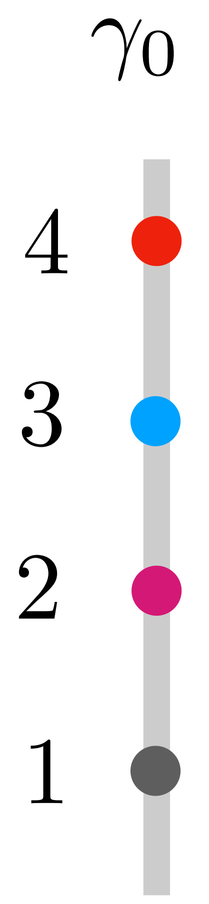
INPUTS
(Independent Variables)

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

CONDITIONAL PROBABILITY $\quad P(y|x)P(x) = P(x|y)P(y)$

BAYES RULE $\quad P(y|x) = \dfrac{P(x|y)P(y)}{P(x)}$

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

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

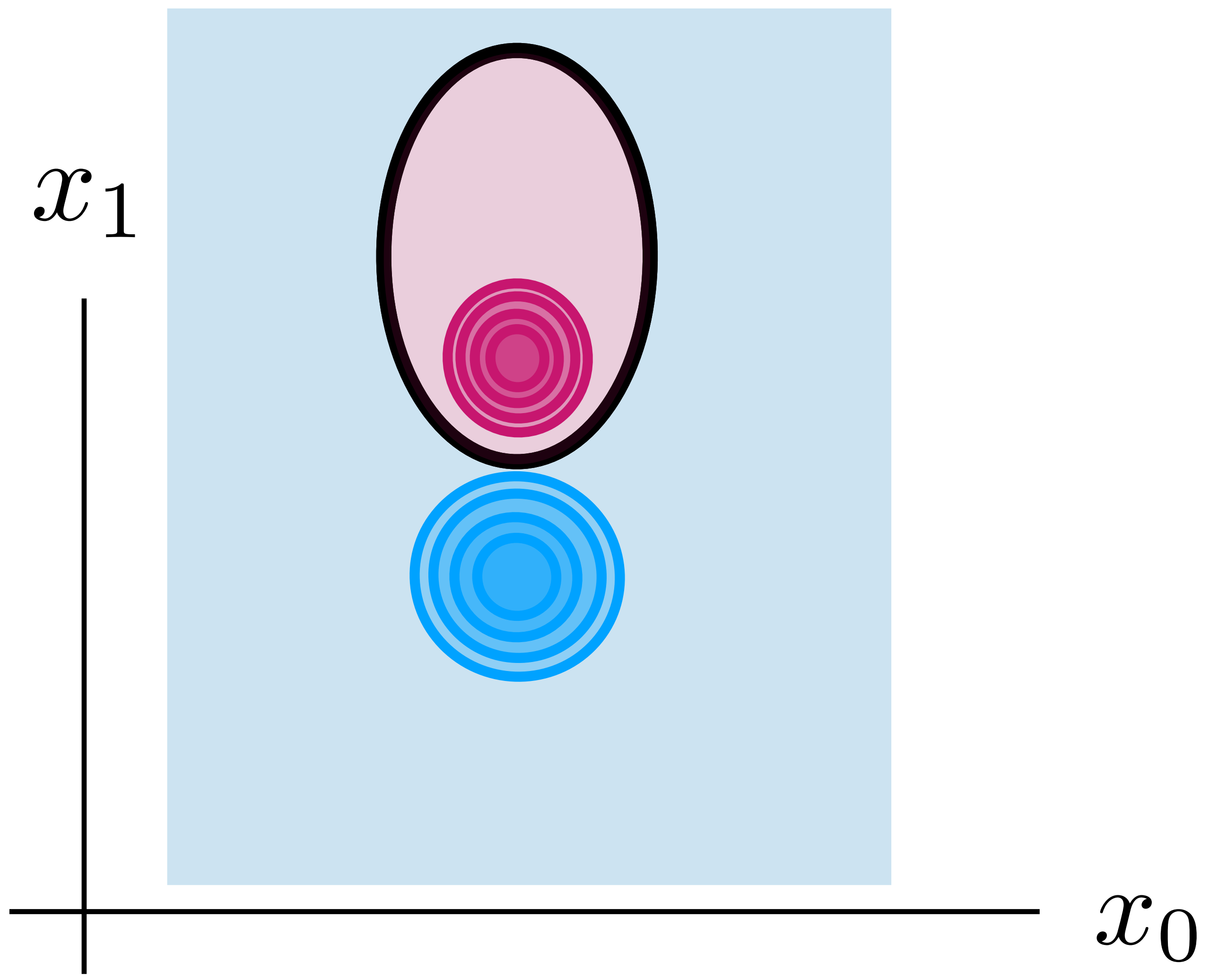
Discrimination Surfaces $\quad \log\ P(y=k|x) = \log\ P(y=k'|x) \qquad$ k vs. k'

Quadratic Discriminant Analysis

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

ellipses or hyperbolas (or parabolas)

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



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

Discriminant Analysis

OUTPUTS
(Dependent Variables)

$$y_t = \theta^T x_t$$

INPUTS
(Independent Variables)

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

CONDITIONAL PROBABILITY

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

BAYES RULE

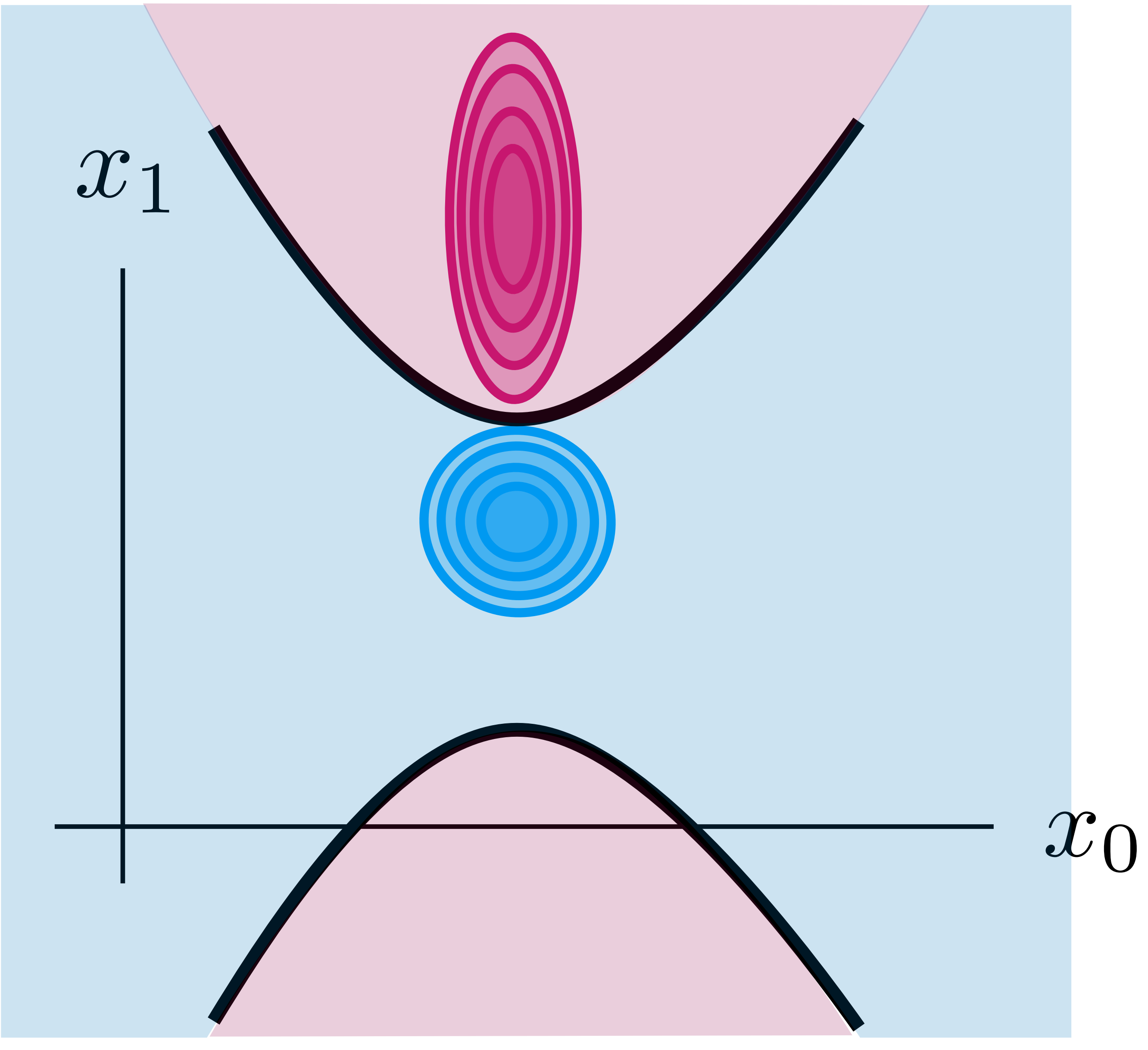
$$P(y|x) = \frac{P(x|y)P(y)}{P(x)}$$

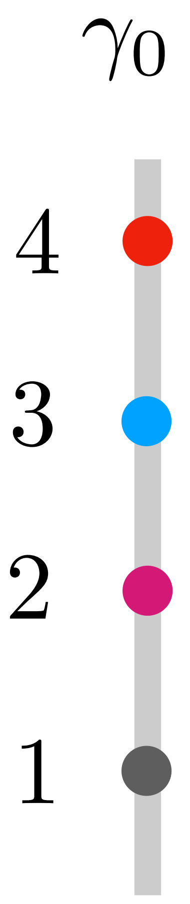
$P(x|y=k)$ ...height of density k

$P(y=k)$ ...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'

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

Quadratic Discriminant Analysis

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

Discriminant Analysis

OUTPUTS (Dependent Variables)

$$y_t = \theta^T x_t$$

INPUTS (Independent Variables)

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

CONDITIONAL PROBABILITY $\quad P(y|x)P(x) = P(x|y)P(y)$

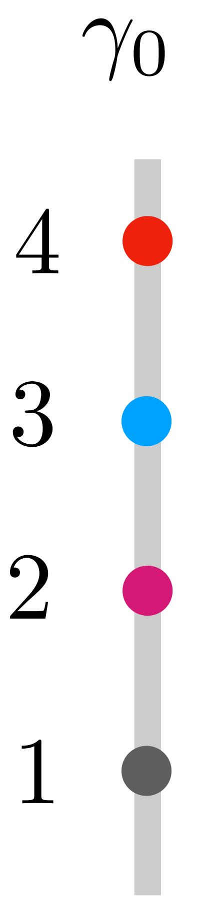
BAYES RULE $\quad P(y|x) = \dfrac{P(x|y)P(y)}{P(x)}$

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

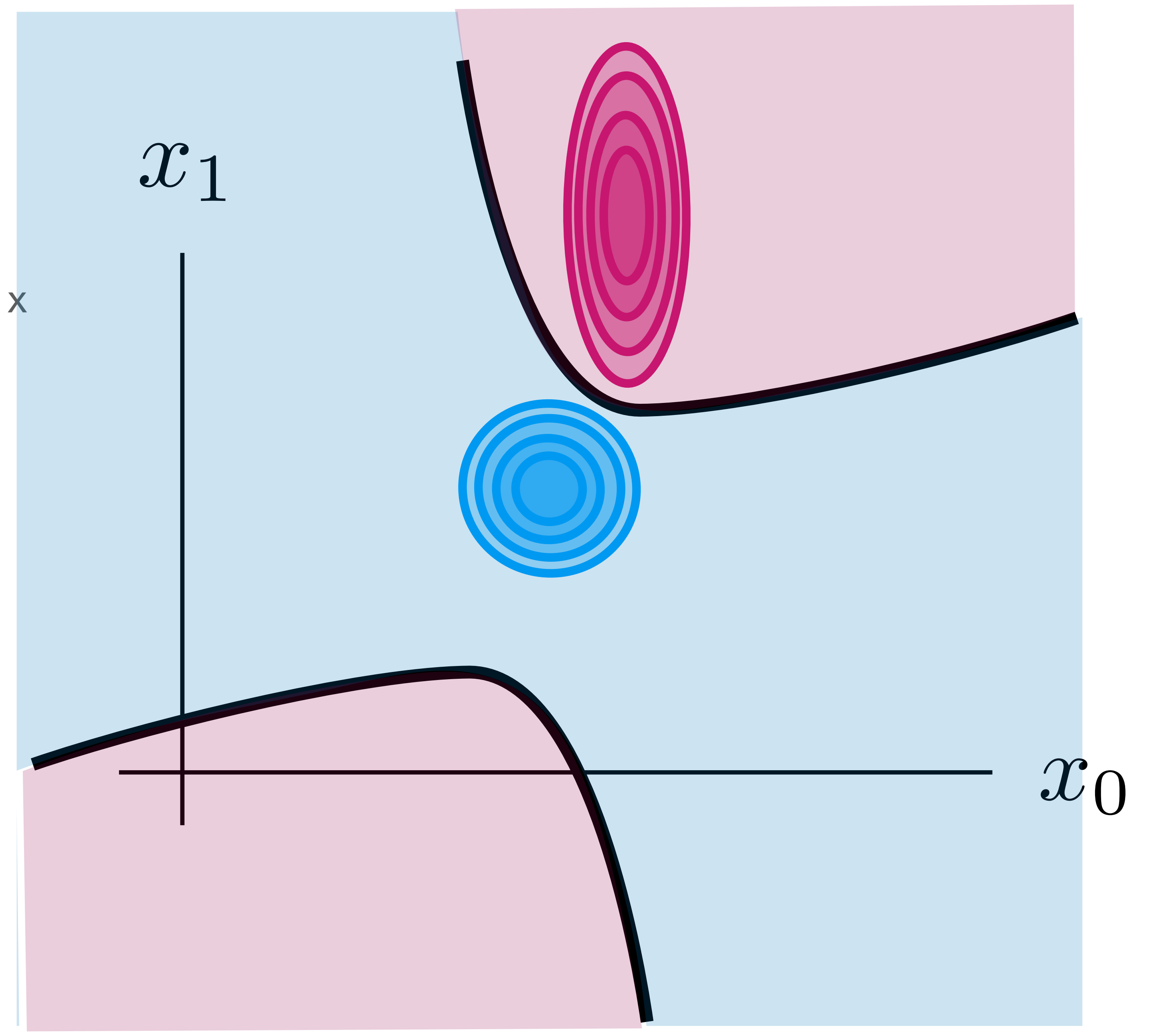
Predicted class $\quad \arg\max_k \ \log\ P(y=k|x)$

Discrimination Surfaces $\quad \log\ P(y=k|x) = \log\ P(y=k'|x) \qquad$ k vs. k’

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

ellipses or hyperbolas (or parabolas)

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



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

Discriminant Analysis

OUTPUTS (Dependent Variables)

$$y_t = \theta^T x_t$$

INPUTS (Independent Variables)

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

CONDITIONAL PROBABILITY

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

BAYES RULE

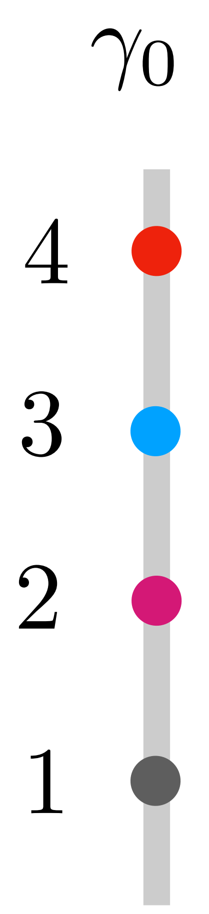
$$P(y|x) = \frac{P(x|y)P(y)}{P(x)}$$

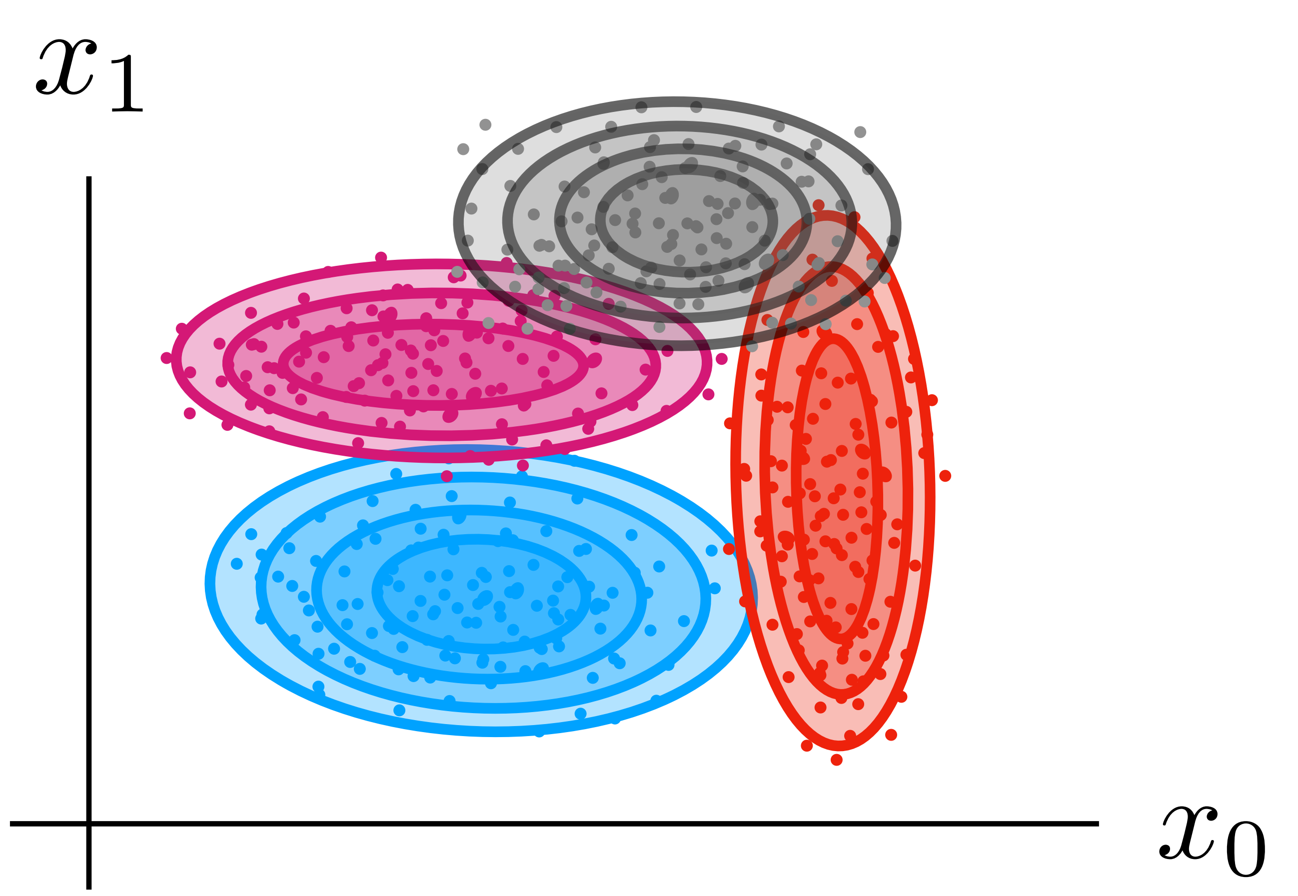
$P(x|y=k)$ ...height of density k

$P(y=k)$ ...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'

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

ellipses or hyperbolas (or parabolas)

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

Gaussian Naive Bayes Classifier

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

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

Discriminant Analysis

OUTPUTS
(Dependent Variables)

$$y_t = \theta^T x_t$$

INPUTS
(Independent Variables)

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

CONDITIONAL PROBABILITY

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

BAYES RULE

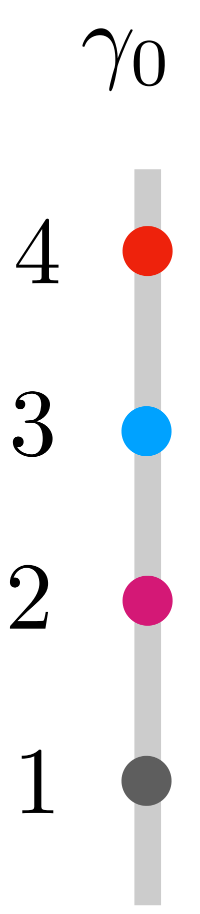
$$P(y|x) = \frac{P(x|y)P(y)}{P(x)}$$

$P(x|y=k)$ …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

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

Mahalanobis Distance

Predicted class

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

Discrimination Surfaces

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

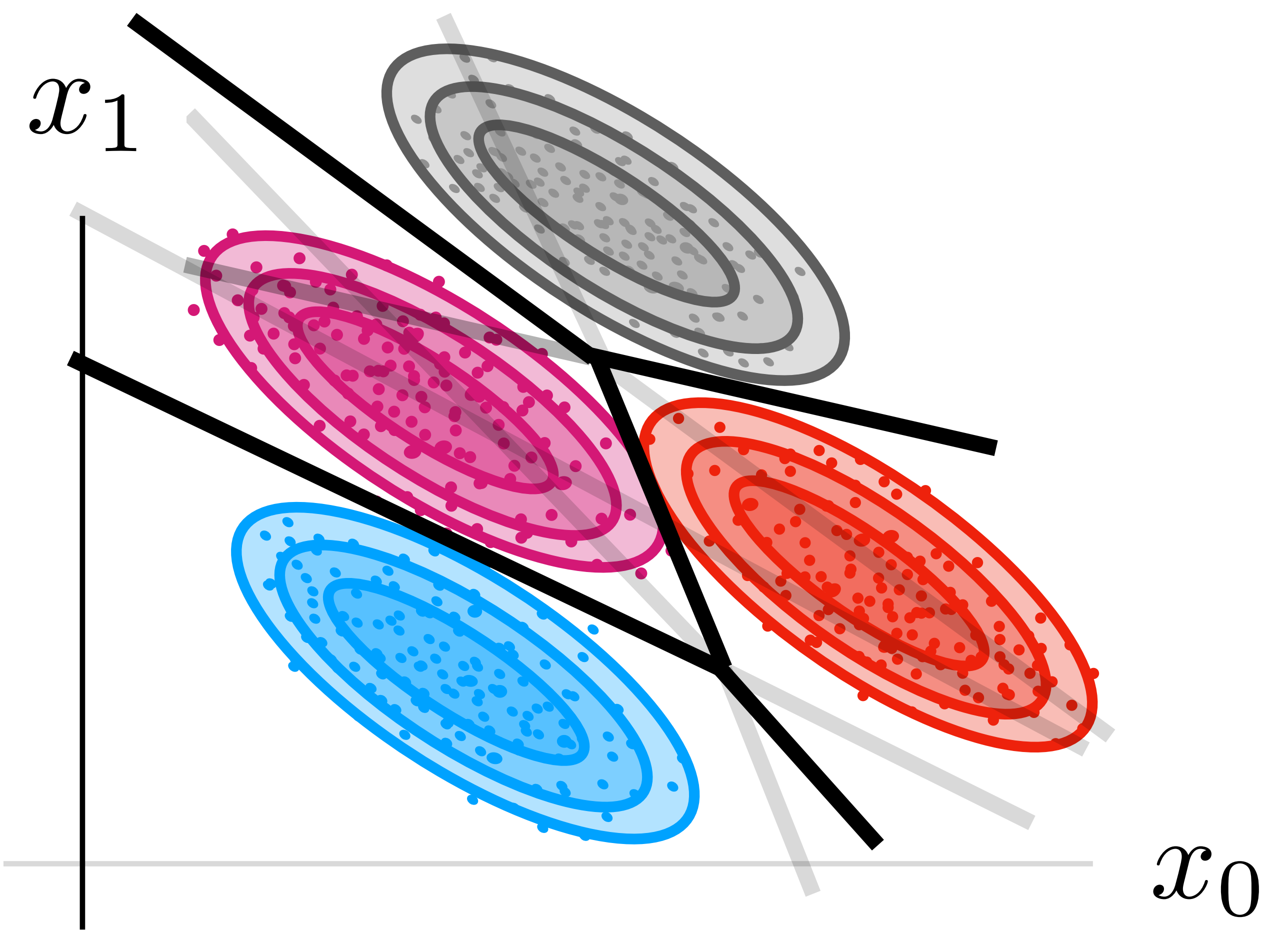
$$\Rightarrow \boxed{\tfrac{1}{2}x^T\cancelto{0}{(\Sigma_k^{-1} - \Sigma_{k'}^{-1})}x - \underbrace{(\mu_k^T\Sigma_k^{-1} - \mu_{k'}^T\Sigma_{k'}^{-1})}x + C_{kk'}}$$

Linear Discriminant Analysis

$$\Rightarrow$$ **hyperplanes**

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

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



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