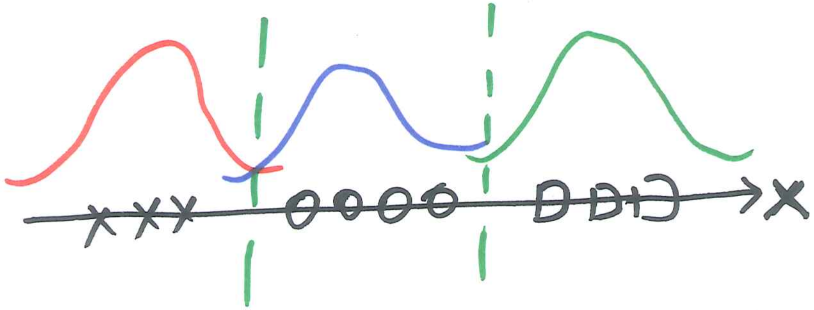
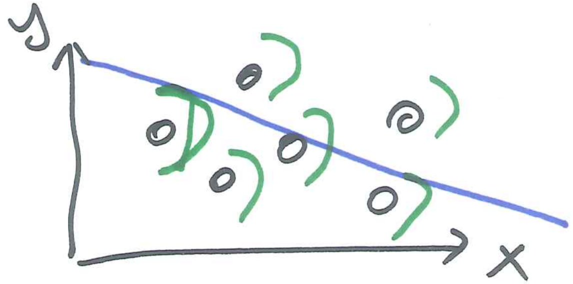
Multivariate Methods





$\Rightarrow$ multiple measurements from our data points

$x_i \in \mathbb{R}^D$

$x_i = [\underbrace{x_{i1}}_{\text{1st feature}}, x_{i2}, x_{i3}, \ldots, \underbrace{x_{iD}}_{\text{Dth feature}}]^T$

$\underbrace{y_i \Rightarrow \text{class labels}}_{\text{classification}}$ $\qquad$ $\underbrace{y_i \Rightarrow \text{target values}}_{\text{regression}}$

$\mathcal{X} = \{(x_i, y_i)\}_{i=1}^{N}$ $\qquad$ $x_i \in \mathbb{R}^D, \; y_i \in \{1, \ldots, K\}$

$y_i \in \mathbb{R}$

$$X = \begin{bmatrix} x_{11}, x_{12}, & \cdots & x_{1D} \\ x_{21} \quad x_{22} & \cdots & x_{2D} \\ \vdots \quad\; \vdots & \ddots & \vdots \\ x_{N1} \quad x_{N2} & \cdots & x_{ND} \end{bmatrix}_{N \times D} \Rightarrow \text{Data matrix} \qquad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

$$\hat{\mu} = \frac{\sum_{i=1}^{N} x_i}{N}$$

$x_i \rightarrow$ D-dimensional vector

column

sample mean

$$\underbrace{\hat{\Sigma}_{D\times D}}_{\text{Sigma}} = \frac{\sum_{i=1}^{N} \underbrace{(x_i-\hat{\mu})}_{D\times 1}\underbrace{(x_i-\hat{\mu})^T}_{1\times D}}{N}$$

sample covariance matrix

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{N} (x_i-\hat{\mu})^2}{N}$$

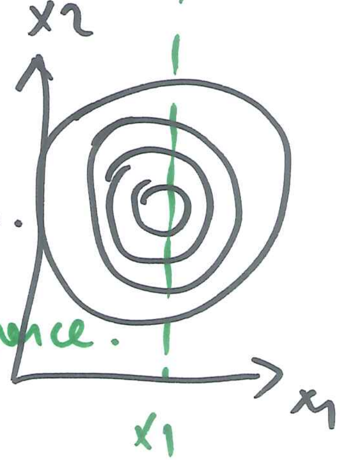
$N > D$

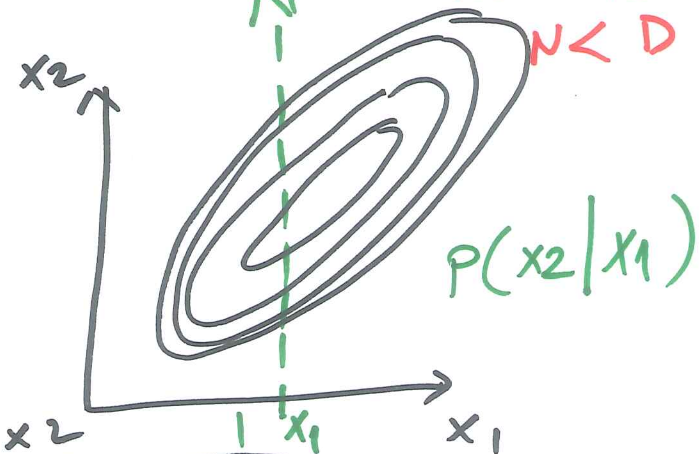
100   10

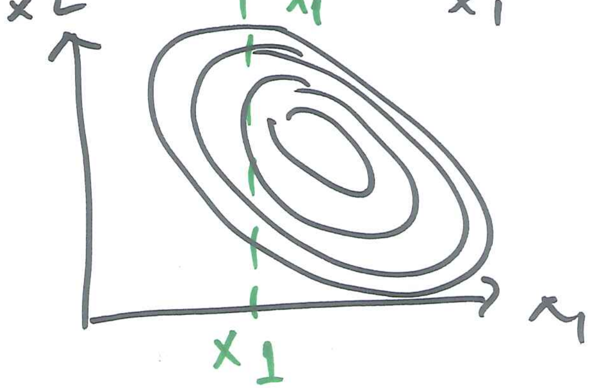
$N < D$

$$\begin{bmatrix} \hat{\sigma}_1^2 & & & \\ & \hat{\sigma}_2^2 & & \\ & & \ddots & \\ & & & \hat{\sigma}_D^2 \end{bmatrix}_{D\times D}$$

covariance terms.







Multivariate Normal Distribution (Gaussian)

univariate $\quad x \sim N(x;\ \underbrace{\mu}_{\text{mean}},\ \underbrace{\sigma^2}_{\text{variance}})$

multivariate $\quad x \sim N(x;\ \underbrace{\mu}_{\text{mean vector}},\ \underbrace{\Sigma}_{\text{covariance matrix}})$

$$N(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$N(x;\mu,\Sigma) = \sqrt{\frac{1}{(2\pi)^D \cdot |\Sigma|}} \cdot \exp\left[-\frac{1}{2}\underbrace{(x-\mu)^T}_{1\times D}\underbrace{\Sigma^{-1}}_{D\times D}\cdot\underbrace{(x-\mu)}_{D\times 1}\right]$$

in 1-D

$$\Sigma = \begin{bmatrix}\sigma_1^2\end{bmatrix}_{1\times 1}$$

$$= \sqrt{\frac{1}{2\pi\sigma^2}} \cdot \exp\left[-\frac{1}{2}\cdot(x-\mu)\frac{1}{\sigma^2}(x-\mu)\right]$$

$$= \sqrt{\frac{1}{2\pi\sigma^2}} \cdot \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

MULTIVARIATE PARAMETRIC CLASSIFICATION

$$P(x|y=c) \sim N(x;\mu_c,\Sigma_c)$$

$$= \sqrt{\frac{1}{(2\pi)^D \cdot |\Sigma_c|}} \cdot \exp\left[-\frac{1}{2}(x-\mu_c)^T\Sigma_c^{-1}\cdot(x-\mu_c)\right]$$

$$\cancel{\sqrt{}}\ (2\pi)^{-D/2}$$

$$\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3$$

$$\hat{\Sigma}_1, \hat{\Sigma}_2, \hat{\Sigma}_3$$

$$g_c(x) = \log\left[p(x|y=c)\,.\,p(y=c)\right]$$

$$= \log \underbrace{p(x|y=c)} + \log p(y=c)$$

$$= -\frac{D}{2}\log(2\pi) - \frac{1}{2}\log(|\hat{\Sigma}_c|) - \frac{1}{2}(x-\hat{\mu}_c)^T.\hat{\Sigma}_c^{-1}.(x-\hat{\mu}_c)$$

$$+ \log \widehat{p(y=c)}$$

$$\hat{\mu}_c = \frac{\sum_{i=1}^{N} 1(y_i=c).x_i}{\sum_{i=1}^{N} 1(y_i=c)}$$ → # of samples in class c

$$\hat{\Sigma}_c = \frac{\sum_{i=1}^{N} 1(y_i=c).(x_i-\hat{\mu}_c)(x_i-\hat{\mu}_c)^T}{\sum_{i=1}^{N} 1(y_i=c)}$$

$$g_c(x) = ax^2 + bx + c$$

$$\widehat{p(y=c)} = \frac{\sum_{i=1}^{N} 1(y_i=c)}{N}$$ → total # of samples

$$g_c\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = [x_1\, x_2]\begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix}.\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + [w_1 \quad w_2]\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + w_0$$

$$= w_{11}x_1^2 + w_{12}x_1.x_2 + w_{21}x_2x_1 + w_{22}x_2^2 + w_1x_1 + w_2x_2 + w_0$$

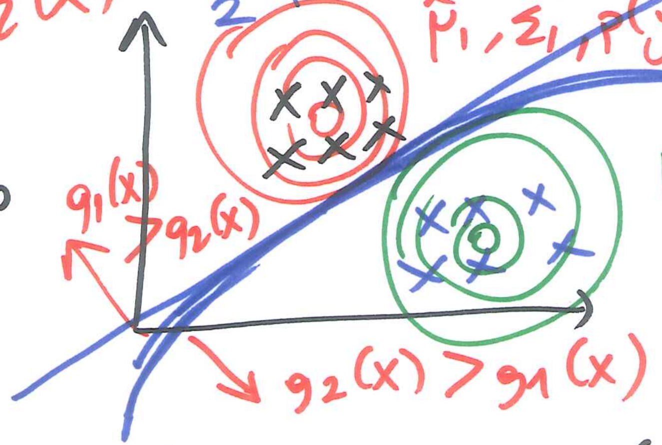
$$g_c(x) = \underset{1\times D}{x^T}\,.\,\underset{D\times D}{W_c}\,.\,\underset{D\times 1}{x} + \underset{1\times D}{w_c^T}\,.\,\underset{D\times 1}{x} + \underset{1\times 1}{w_{c0}}$$

$$W_c = -\frac{1}{2}\hat{\Sigma}_c^{-1} \qquad w_c = \hat{\Sigma}_c^{-1}\cdot\hat{\mu}_c \qquad w_{c0} = -\frac{1}{2}\hat{\mu}_c^T\cdot\hat{\Sigma}_c^{-1}\cdot\hat{\mu}_c - \frac{1}{2}\log|\hat{\Sigma}_c| + \log \widehat{P(y=c)}$$

$$\left.\begin{matrix} g_1(x) \\ g_2(x) \end{matrix}\right\} \; g_1(x) = g_2(x)$$

$$g_1(x) = x^T W_1 x + w_1^T x + w_{10}$$

$$g_2(x) = x^T W_2 x + w_2^T x + w_{20}$$



$$g_1(x) = g_2(x)$$

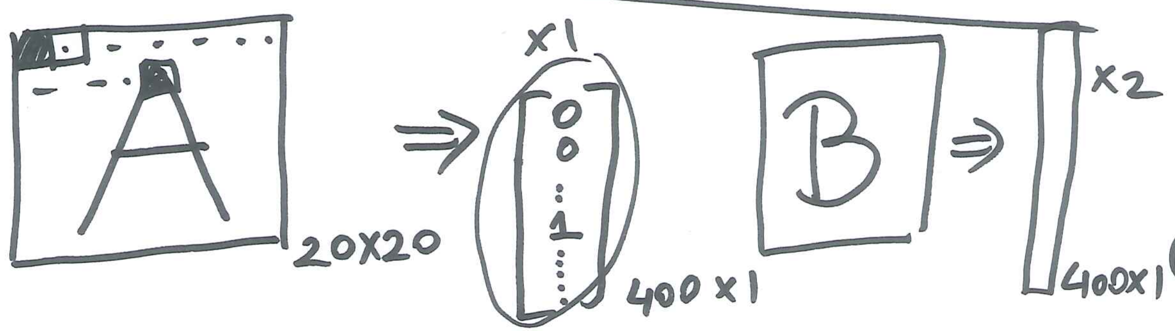
$$\left.\begin{matrix} g_1(x) = a_1x^2 + b_1x + c_1 \\ g_2(x) = a_2x^2 + b_2x + c_2 \end{matrix}\right\rangle \; (a_1 - a_2)x^2 + \underbrace{(b_1 - b_2)}_{0}x + (c_1 - c_2) = 0$$

$$x^2 - 9 = 0 \Rightarrow x = \pm 3$$

$$x^2 - 8x + 16 = 0 \Rightarrow (x-4)^2 = 0$$

$$\underline{\underline{x = +4}}$$

$$X = \begin{array}{c} \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1D} \\ x_{21} & x_{22} & \cdots & x_{2D} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{ND} \end{bmatrix} \\ N \times D \end{array} \begin{array}{l} \rightarrow x_1 \\ \rightarrow x_2 \\ \vdots \\ \rightarrow x_N \end{array} \Rightarrow \text{discrete.}$$



$$p(x \mid y=c) = \prod_{d=1}^{D} p_{cd}^{\,x_d} (1-p_{cd})^{1-x_d}$$

$\hookrightarrow$ probability of having a "black" pixel at position #d $\Big\}$ p_{cd}

$1-p_{cd}$

$$g_c(x) = \log p(x \mid y=c) + \log p(y=c)$$

$$g_c(x) = \sum_{d=1}^{D} \Big[x_d \cdot \log(p_{cd}) + (1-x_d) \cdot \log(1-p_{cd}) \Big] + \log \boxed{p(y=c)}$$

$$\hat{p}_{cd} = \frac{\sum_{i=1}^{N} x_{id}\, \mathbb{1}(y_i=c)}{\sum_{i=1}^{N} \mathbb{1}(y_i=c)}$$

$$\hat{p}(y=c) = \frac{\sum_{i=1}^{N} \mathbb{1}(y_i=c)}{N}$$

$$\log 0 \equiv 0$$

$$\underline{g_A(x)},\ \underline{g_B(x)},\ \ldots,\ \underline{g_Z(x)} \qquad \boxed{?} \rightarrow C$$