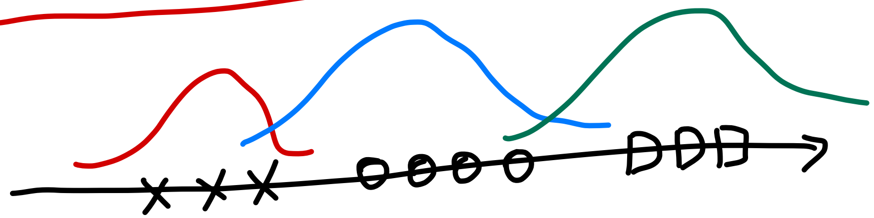
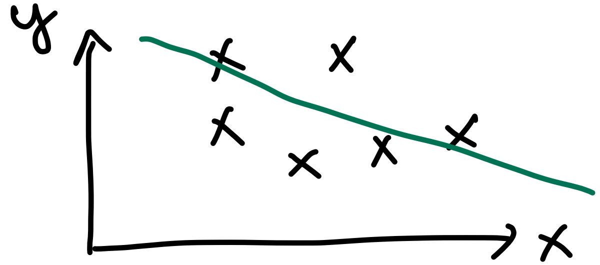
Multivariate Methods





$\Rightarrow$ multiple measurements from our data points

$x_i \in \mathbb{R}^D \qquad \mathbf{x}_i = [x_{i1} \quad x_{i2} \quad \cdots\cdots \quad x_{iD}]^T$

ith data point

$\hookrightarrow$ first feature

$\hookrightarrow$ D^{th} feature

$y_i \Rightarrow$ class labels

classification

$y_i \Rightarrow$ target values

regression

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

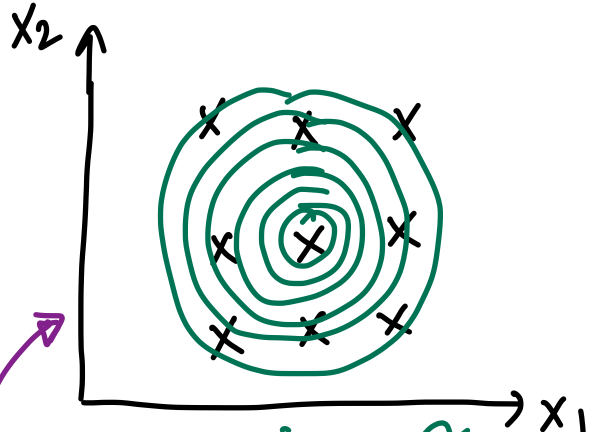
$$y_i \in \{1, 2, \ldots, K\}$$

$$y_i \in \mathbb{R}$$

$$X = \underbrace{\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1D} \\ x_{21} & x_{22} & \cdots & x_{2D} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{ND} \end{bmatrix}}_{N \times D} \begin{array}{l} \to x_1 \\ \to x_2 \\ \vdots \\ \to x_N \end{array}$$

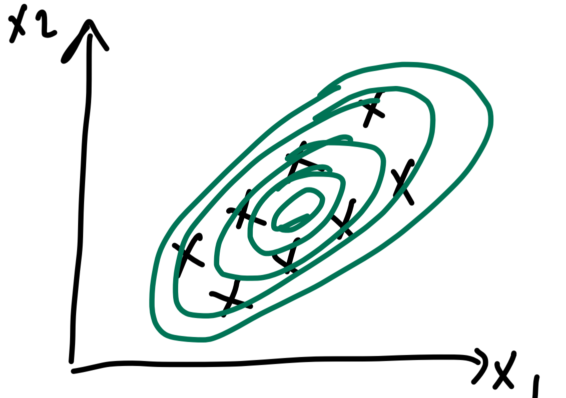
data matrix

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}_{N \times 1}$$



covariance $\simeq 0$

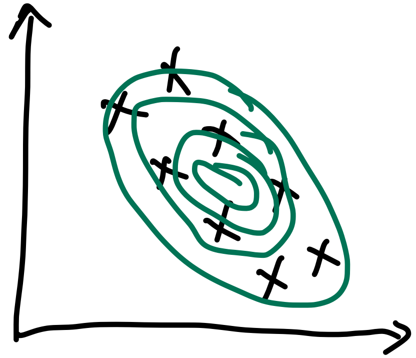
$$\begin{bmatrix}\sigma_1^2 & 0 \\ 0 & \sigma_2^2\end{bmatrix}_{2\times 2}$$



Covariance > 0

$$\begin{bmatrix}\sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2\end{bmatrix}_{2\times 2}$$

$\sigma_{12} = \sigma_{21} > 0$



Covariance < 0

$$\begin{bmatrix}\sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2\end{bmatrix}_{2\times 2}$$

$\sigma_{12} = \sigma_{21} < 0$

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

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

$\hat{\mu}$ = sample mean vector

$\hat{\Sigma}$ = sample covariance matrix

sample mean

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

sample variance

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

$$\hat{\mu}_{D\times 1} = \frac{\sum_{i=1}^{N} \underbrace{x_i}_{D\times 1}}{N} \quad \text{vector}$$

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

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

Σ has to be invertible

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

when $D=1$

$\Sigma = [\sigma^2]_{1\times 1}$

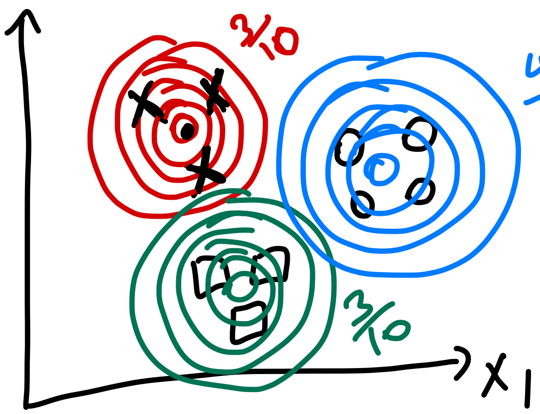
$$= \frac{1}{\sqrt{(2\pi)^1 \cdot \sigma^2}} \cdot \exp\left[-\frac{1}{2}(x-\mu)\boxed{\frac{1}{\sigma^2}}(x-\mu)\right]$$

Σ^{-1}

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \Rightarrow \text{univariate Gaussian density.}$$

Multivariate Parametric Classification

D=2 ⇒



$$p(\mathbf{x}|y=c) \sim N(\mathbf{x}; \mu_c, \Sigma_c)$$

⇓

Class conditional density

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

Model parameters

$D \Leftarrow \hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3 \Rightarrow D$

$\hat{\Sigma}_1, \hat{\Sigma}_2, \hat{\Sigma}_3 \Rightarrow \frac{D(D+1)}{2}$

$\hat{P}(y=1), \hat{P}(y=2), \hat{P}(y=3) \Rightarrow \overbrace{K-1}^{\text{in total}}$

# of parameters $= K.D + K\frac{D.(D+1)}{2} + K-1$

$$g_c(\mathbf{x}) = \log\left[p(\mathbf{x}|y=c)\right] + \log\left[P(y=c)\right]$$

$$= -\frac{D}{2}.\log(2\pi) - \frac{1}{2}\log(|\hat{\Sigma}_c|) - \frac{1}{2}.(\mathbf{x}-\hat{\mu}_c)^T.\hat{\Sigma}_c^{-1}.(\mathbf{x}-\hat{\mu}_c) + \log\left[\hat{P}(y=c)\right]$$

$(x-a).b(x-a) \Rightarrow bx^2 - 2abx + ba^2$

We have to find $\hat{\mu}_1, \ldots, \hat{\mu}_K, \hat{\Sigma}_1, \ldots, \hat{\Sigma}_K, \hat{P}(y=1), \hat{P}(y=2), \ldots, \hat{P}(y=K)$ from our training set.

$$\hat{\mu}_c = \frac{\sum_{i=1}^{N} [x_i \cdot 1(y_i = c)]}{N_c}$$

$N_c = \sum_{i=1}^{N} 1(y_i = c)$

# of data points from class #c.

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

$$\boxed{a^T \cdot b = b^T \cdot a}$$

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

$\Rightarrow$ frequency of class #c

$$\underbrace{g_c(x)}_{\text{scalar}} = \underbrace{\underbrace{x^T}_{1\times D} \cdot \underbrace{W_c}_{D\times D} \underbrace{x}_{D\times 1}}_{\text{scalar}} + \underbrace{\underbrace{w_c^T}_{1\times D} \cdot \underbrace{x}_{D\times 1}}_{\text{scalar}} + \underbrace{\underbrace{w_{c0}}_{1\times 1}}_{\text{scalar}}$$

$x^T \cdot w_c$

$W_c = ?$

$w_c = ?$

$w_{c0} = ?$

$$-\frac{1}{2}(x - \hat{\mu}_c)^T \cdot \hat{\Sigma}_c^{-1} \cdot (x - \hat{\mu}_c) = -\frac{1}{2} x^T \hat{\Sigma}_c^{-1} \cdot x + x^T \cdot \hat{\Sigma}_c^{-1} \cdot \hat{\mu}_c - \frac{1}{2} \hat{\mu}_c^T \hat{\Sigma}_c^{-1} \cdot \hat{\mu}_c$$

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

$$\left.\begin{aligned}
g_1(x) &= x^T.W_1.x + w_1^T.x + w_{10} \\
g_2(x) &= x^T.W_2.x + w_2^T.x + w_{20} \\
&\;\vdots \\
g_K(x) &= x^T.W_K.x + w_K^T.x + w_{K0}
\end{aligned}\right\}$$ pick the maximum one.

---

when $K=2$

$$\left.\begin{aligned}
12 \quad 1002 \quad g_1(x) &= x^T.W_1.x + w_1^T.x + w_{10} \\
11 \quad 1001 \quad g_2(x) &= x^T.W_2.x + w_2^T.x + w_{20}
\end{aligned}\right\}$$ pick the maximum one

$$g_1(x) - g_2(x) = x^T\underbrace{(W_1 - W_2)}_{W}.x + \underbrace{(w_1 - w_2)^T}_{w^T}.x + \underbrace{(w_{10} - w_{20})}_{w_0}$$

$$g(x) = x^T W x + w^T.x + w_0$$

if $g(x) > 0 \Rightarrow$ pick the first class

$< 0 \Rightarrow$ pick the second class

$$\hat{\Sigma}_c = \frac{\sum_{i=1}^{N} \left[(x_i - \hat{\mu}_c)(x_i - \hat{\mu}_c)^T \, 1(y_i = c) \right]}{N_c}$$

when $D = 10$
$N_c = 8$
$\Downarrow$
problem = ?

$$[\;][\;]^T + [\;][\;]^T + \dots + [\;][\;]^T$$

$\Downarrow$ the resulting covariance matrix has rank $\min(D, N_c)$

rank-deficient
rank-one matrix

$(x_i - \hat{\mu}_c)$ $\quad (x_i - \mu_c)^T$

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \qquad \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix} \Rightarrow \text{this is invertible.}$$