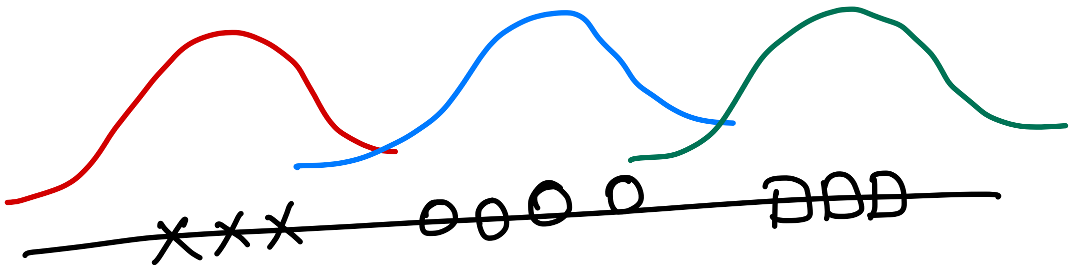
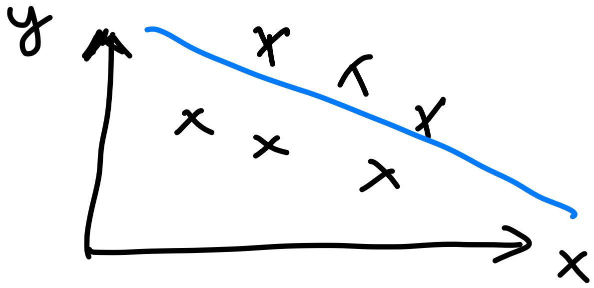
Multivariate Methods





$\Rightarrow$ multiple measurements from our data points

$x_i \in \mathbb{R}^D$ ith data point

$x_i = [\,x_{i1} \quad x_{i2} \quad - - - - - \quad x_{iD}\,]^T$

x_{i1} → first feature

x_{i2} → second feature

x_{iD} → D^{th} feature

$y_i \Rightarrow$ class labels: classification

$y_i \Rightarrow$ target values: regression

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

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

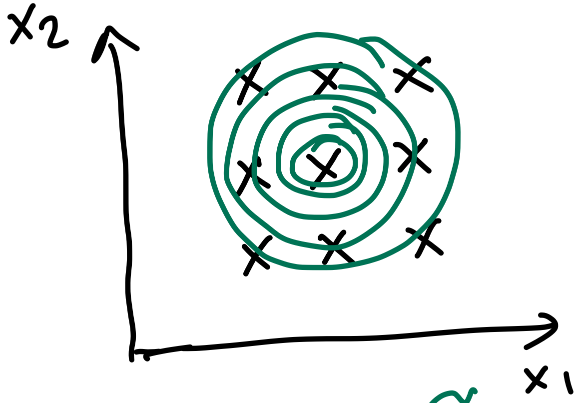
$$y_i \in \mathbb{R}$$

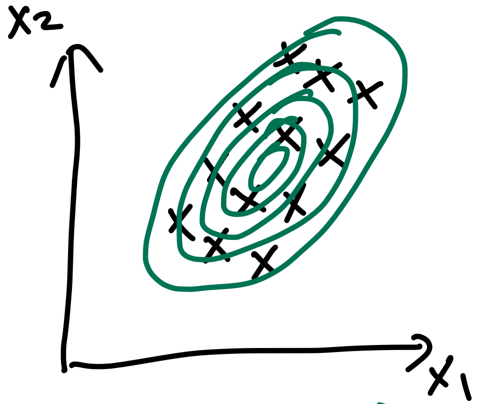
data matrix →

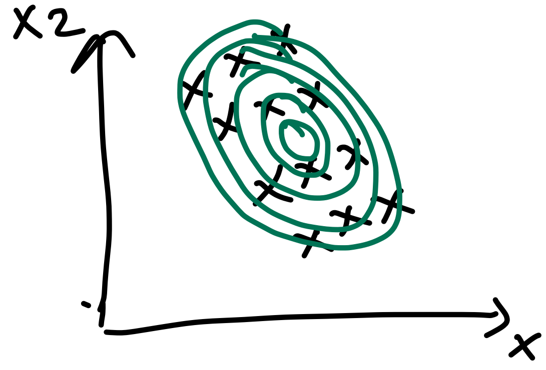
$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1D} \\ x_{21} & x_{22} & \cdots & x_{2D} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{ND} \end{bmatrix}_{N \times D} \begin{array}{l} \rightarrow x_1 \\ \rightarrow x_2 \\ \quad \vdots \\ \rightarrow x_N \end{array}$$

label (or output) vector →

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







covariance $\cong 0$

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

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

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

sample mean vector

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

$x_i \rightarrow$ D-dimensional column vectors

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

sample covariance matrix

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

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

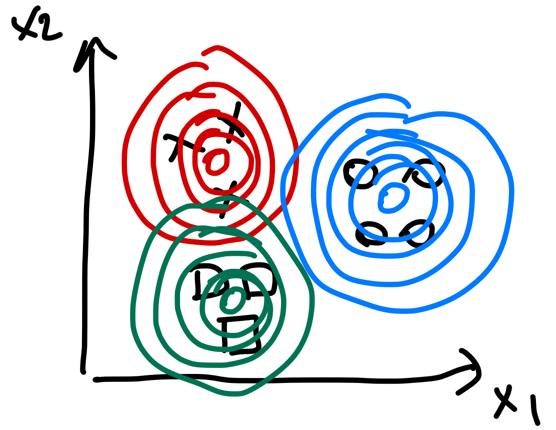
when $D=1 \Rightarrow$

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

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

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

Multivariate Parametric Classification



when $D = 2$

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

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

model parameters

$$\hat{\mu}_1 \quad \hat{\mu}_2 \quad \hat{\mu}_3$$
$$\hat{\Sigma}_1 \quad \hat{\Sigma}_2 \quad \hat{\Sigma}_3$$
$$\underbrace{\hat{P}(y=1) \quad \hat{P}(y=2) \quad \hat{P}(y=3)}$$

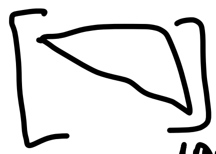
We need to estimate these from data.

$$\underbrace{g_c(x)}_{\text{score for class\#c}} = \log\left[p(x \mid y=c) \cdot P(y=c)\right]$$

$$= \underbrace{\log\left[p(x \mid y=c)\right]}_{} + \underbrace{\log\left[P(y=c)\right]}_{}$$

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

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



1000×1000 $\rightarrow \dfrac{1000 \times 1001}{2}$

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

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

↳ # of data points from class # c.

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

$N_c = 10 \quad D = 20$

Is $\hat{\Sigma}_c$ invertible?

$$\hat{P}(y=c) = \frac{\sum_{i=1}^{N} 1(y_i=c)}{N} = \frac{N_c}{N} \Rightarrow \text{frequency of class } \# c$$

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

$w_c^T \cdot x = x^T \cdot w_c$

$W_c = ? \; -\frac{1}{2} \hat{\Sigma}_c^{-1}$

$w_c = ? \; \hat{\Sigma}_c^{-1} \cdot \hat{\mu}_c$

$w_{c0} = ? \; + \; \underline{\qquad}$ from previous page

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

$$\underbrace{-\frac{1}{2} x^T \cdot \hat{\Sigma}_c^{-1} \cdot x}_{\text{quadratic}} + \underbrace{x^T \cdot \hat{\Sigma}_c^{-1} \cdot \hat{\mu}_c}_{\text{linear}} \underbrace{- \frac{1}{2} \hat{\mu}_c^T \hat{\Sigma}_c^{-1} \cdot \hat{\mu}_c}_{\text{constant}}$$

$$W_c = -\frac{1}{2}\hat{\Sigma}_c^{-1}$$

$$w_c = \hat{\Sigma}_c^{-1}\cdot\hat{\mu}_c$$

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

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

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

$$\vdots$$

$$g_K(x) = x^T.W_K\,x + w_K^T.x + w_{K0}$$

pick the maximum one.

---

when $K=2$

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

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

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

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

if $g(x) < 0 \Rightarrow$ assign to the 2nd class

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

$$\operatorname{rank}(\underset{4\times 4,\ 3}{A}+\underset{4\times 4,\ 3}{B}) \le \operatorname{rank}(A)+\operatorname{rank}(B)$$

$$\underbrace{\begin{bmatrix} \\ \\ \\ \end{bmatrix}_{D\times 1}\begin{bmatrix} \quad \end{bmatrix}_{1\times D}}_{\text{rank-one matrix}} + \cdots + \underbrace{\begin{bmatrix} \\ \\ \\ \end{bmatrix}_{D\times 1}\begin{bmatrix} \quad \end{bmatrix}_{1\times D}}_{\text{rank-one matrix}} =$$

$$\underbrace{\begin{bmatrix} 1 \\ 2 \end{bmatrix}\begin{bmatrix} 1 & 2 \end{bmatrix}}_{\text{rank-one}} + \underbrace{\begin{bmatrix} 20 \\ 41 \end{bmatrix}\begin{bmatrix} 20 & 41 \end{bmatrix}}_{\text{rank-one}}$$

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} + \begin{bmatrix} 0 & 4 & 0 & 8 \\ 0 & 8 & 1 & 16 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 5 & 2 & 10 \\ 2 & 10 & 5 & 20 \end{bmatrix}}_{\text{rank-two}}$$