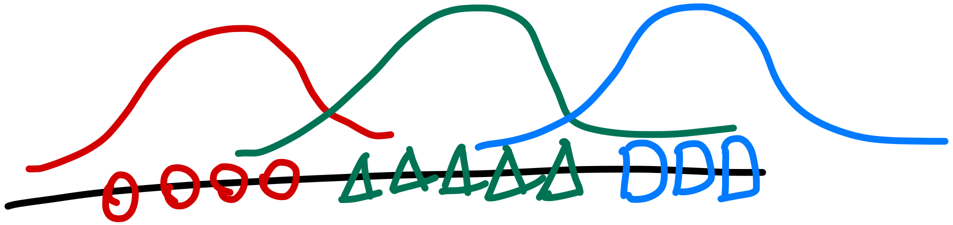
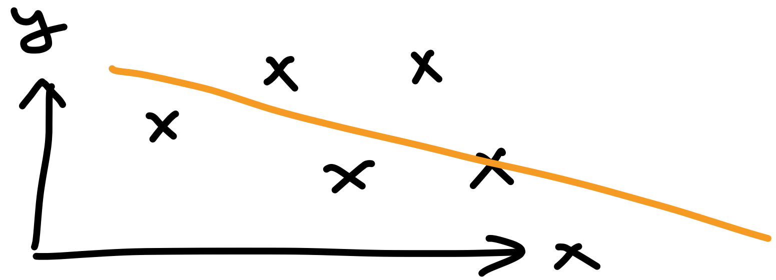
Multivariate methods





$\Rightarrow$ multiple measurements are taken from each point

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

i^{th} data point $\qquad$ ↳ first feature $\qquad$ → second feature $\qquad$ ↳ D^{th} feature

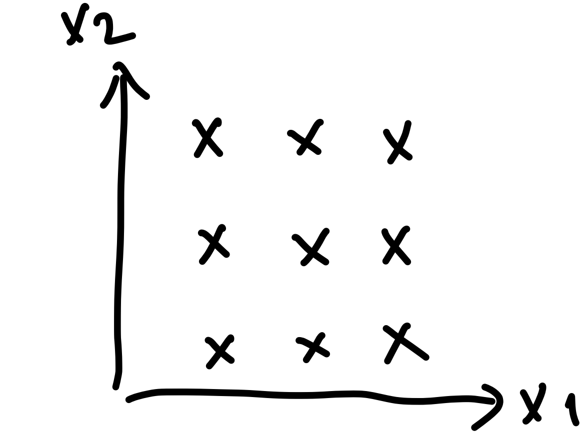
$y_i \Rightarrow$ class label $\qquad$ $y_i \Rightarrow$ target value.

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

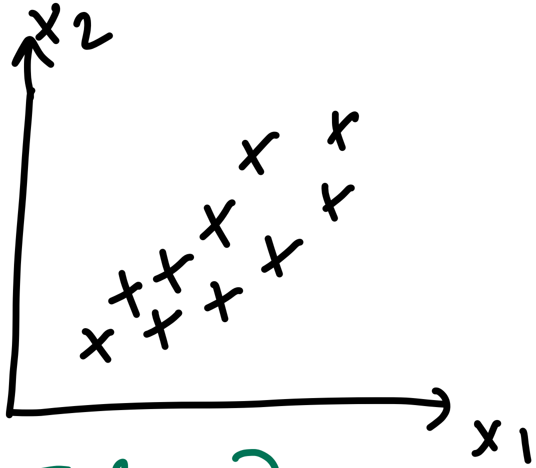
$y_i \in \{1, 2, ---, K\}$

$y_i \in \mathbb{R}^1$

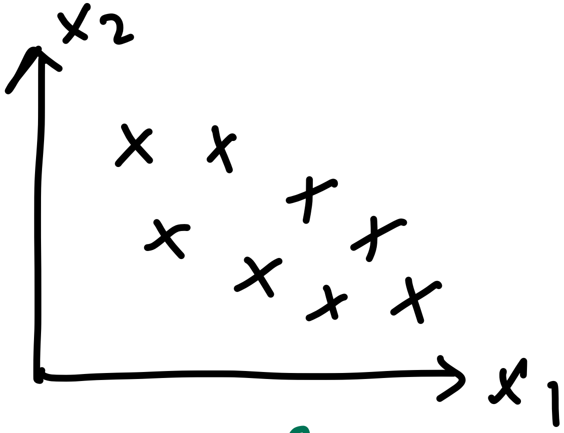
$$\mathbf{X} = \underbrace{\begin{bmatrix} x_{11} & x_{12} & --- & x_{1D} \\ x_{21} & x_{22} & --- & x_{2D} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & --- & x_{ND} \end{bmatrix}}_{N \times D} \begin{matrix} \rightarrow \mathbf{x}_1^T \\ \rightarrow \mathbf{x}_2^T \\ \\ \rightarrow \mathbf{x}_N^T \end{matrix} \qquad \mathbf{y} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}}_{N \times 1}$$



$$\begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} \Rightarrow \begin{matrix} \sigma_{12} \cong 0 \\ \sigma_{21} \cong 0 \end{matrix}$$



$$\begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} \Rightarrow \sigma_{12} = \sigma_{21} > 0$$



$$\begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} \Rightarrow \sigma_{12} = \sigma_{21} < 0$$

Univariate $\Rightarrow x \sim N(x; \mu, \sigma^2)$

multivariate $\Rightarrow \mathbf{x} \sim N(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$

mean vector

covariance matrix

$$\hat{\mu}_{1\times1} = \frac{\sum_{i=1}^{N} x_{i\,1\times1}}{\,}$$

$$\hat{\sigma}^2_{1\times1} = \frac{\sum_{i=1}^{N} (x_i - \hat{\mu})^2_{1\times1}}{N}$$

$$\hat{\boldsymbol{\mu}}_{D\times1} = \frac{\sum_{i=1}^{N} \mathbf{x}_{i\,D\times1}}{N}$$

$$\hat{\boldsymbol{\Sigma}}_{D\times D} = \frac{\sum_{i=1}^{N} (\mathbf{x}_i - \hat{\boldsymbol{\mu}})_{D\times1} (\mathbf{x}_i - \hat{\boldsymbol{\mu}})^T_{1\times D}}{N}$$

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

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

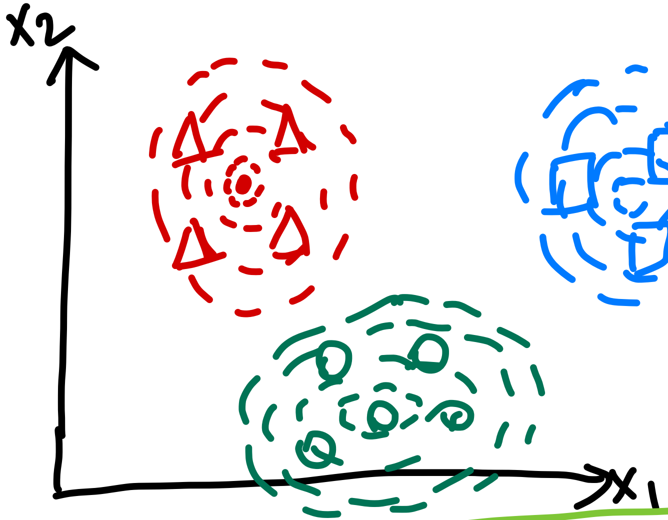
when $D=1$

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

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

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

Multivariate Parametric Classification



class conditional density

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

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

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

$$\boxed{-\frac{1}{2}(x-a)^T.B(x-a) = -\frac{1}{2}x^T.Bx + x^T.Ba - \frac{1}{2}a^T.Ba}$$

$a^T.Bx$

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

$$-\frac{1}{2}x^T.\hat{\Sigma}_c^{-1}.x + x^T.\hat{\Sigma}_c^{-1}.\hat{\mu}_c - \frac{1}{2}\hat{\mu}_c^T\hat{\Sigma}_c^{-1}\hat{\mu}_c$$

total # of parameters

$$= \underbrace{K.D}_{\text{mean vectors}} + \underbrace{K.\left[\frac{D.(D+1)}{2}\right]}_{\text{covariance matrices}} + \underbrace{K-1}_{\text{prior probabilities}}$$

$$(A.b)^T = b^T.A^T$$

$$\boxed{a^T.Xb = b^T.Xa}$$
if X is square & symmetric

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

$$\hat{\mu}_c = \frac{\sum_{i=1}^{N}\left[x_i \, 1(y_i = c)\right]}{\sum_{i=1}^{N} 1(y_i = c)}$$ ← first order.

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

← constant

$$\hat{Pr}(y = c) = \frac{\sum_{i=1}^{N} 1(y_i = c)}{N} = \frac{N_c}{N}$$ ⇒ frequency of class c.

$W_c = ?$

$w_c = ?$

$w_{c0} = ?$

$$g_c(x) = x^T . W_c x + w_c^T . x + w_{c0}$$

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

w_c: $\hat{\mu}_c^T . \hat{\Sigma}_c^{-1}$

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

$$g_1(x) = x^T . W_1 . x + w_1^T . x + w_{10}$$
$$g_2(x) = x^T . W_2 . x + w_2^T . x + w_{20}$$
$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots$$
$$g_K(x) = x^T . W_K x + w_K^T . x + w_{K0}$$

Pick the maximum one.

---

when $K=2$

$$g_1(x) = x^T . W_1 . x + w_1^T . x + w_{10}$$
$$g_2(x) = x^T . W_2 . x + w_2^T . x + w_{20}$$

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

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

$\rightarrow > 0 \Rightarrow$ 1st class

$\rightarrow < 0 \Rightarrow$ 2nd class