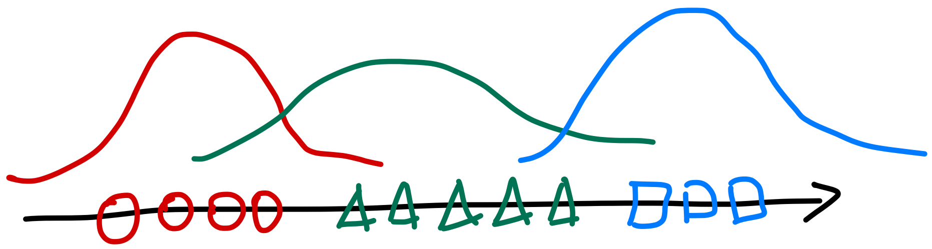
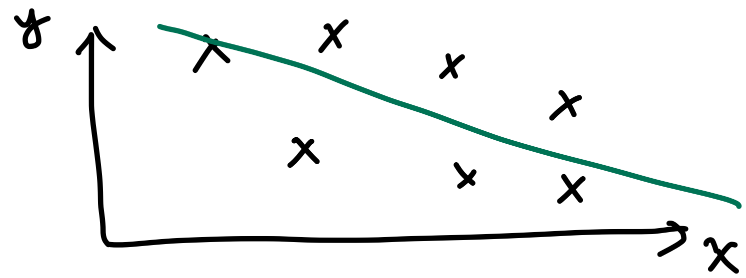
Multivariate methods





$\Rightarrow$ multiple measurements from each data point

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

ith data point; first feature; D^{th} feature

$y_i \Rightarrow$ class label
classification

$y_i \Rightarrow$ target value
regression

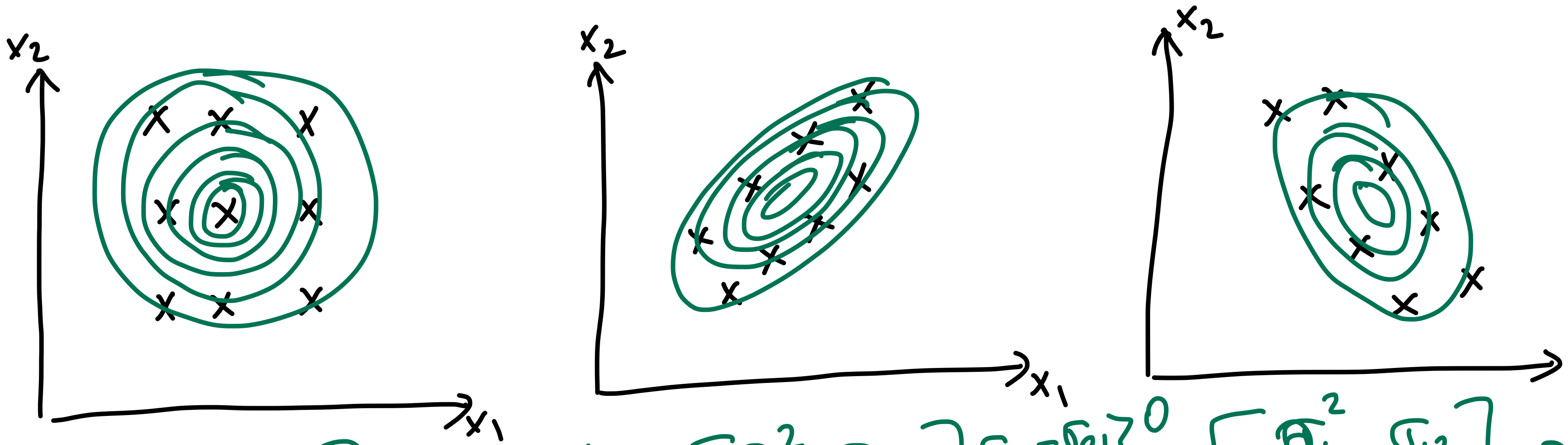
$$\mathcal{X} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N} \qquad \mathbf{x}_i \in \mathbb{R}^D$$

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

$y_i \in \mathbb{R}$

$$\mathbf{X} = \begin{array}{c} \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1D} \\ x_{21} & x_{22} & \cdots & x_{2D} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{ND} \end{bmatrix}_{N \times D} \end{array} \begin{array}{l} \rightarrow \mathbf{x}_1 \\ \rightarrow \mathbf{x}_2 \\ \\ \rightarrow \mathbf{x}_N \end{array}$$

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



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

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

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

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

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

↳ covariance matrix (D×D)

↳ mean vector (D×1)

sample mean

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

sample mean vector

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

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

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

sample variance

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

sample covariance matrix

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

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

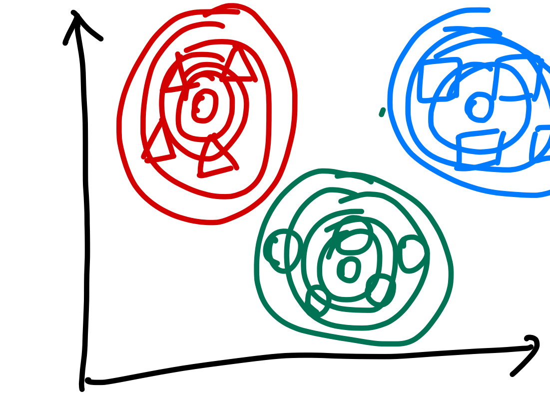
when $D=1$

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

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

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

Multivariate Parametric Classification



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

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

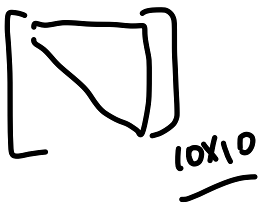
$$g_c(x) = \log\left[p(x \mid y=c)\right] + \log\left[\Pr(y=c)\right]$$

$$= -\frac{D}{2}\cdot\log(2\pi) - \frac{1}{2}\log(|\hat{\Sigma}_c|)$$

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

Model parameters

$\hat{\mu}_1$ (D×1)	$\hat{\mu}_2 \rightarrow D\times 1$	$\hat{\mu}_3 \rightarrow D\times 1$
$\hat{\Sigma}_1 \rightarrow D\times D$	$\hat{\Sigma}_2 \rightarrow D\times D$	$\hat{\Sigma}_3 \rightarrow D\times D$
$\hat{\Pr}(y=1) \rightarrow 1\times 1$	$\hat{\Pr}(y=2) \rightarrow 1\times 1$	$\hat{\Pr}(y=3) \rightarrow 1\times 1$



$$-\frac{1}{2}(x-a)^T B(x-a)$$

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

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

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

$$\boxed{a^T.X.b = b^T.X\,a \quad \text{if } X \text{ is square \& symmetric}}$$

total # of parameters

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

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

N_c = # of data points from class c.

$$= a.x^2 + bx + c$$

$$= x^T A . x + b^T . x + c$$

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

$$\boxed{(A^T . B)^T = B^T . A}$$

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

frequency of class c.

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

$$w_c = ? \quad \hat{\Sigma}_c^{-1} . \hat{\mu}_c$$

$$w_{c0} = ?$$

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

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

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

$$g_1(x) = x^T W_1 . x + w_1^T . x + w_{10}$$
$$g_2(x) = x^T . W_2 . x + w_2^T . x + w_{20}$$
$$\vdots \qquad \vdots$$
$$g_k(x) = x^T . W_k . x + w_k^T . x + w_{k0}$$

Pick the maximum one.

---

when $k = 2$

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

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

$g(x) > 0 \Rightarrow$ pick 1st class

$g(x) < 0 \Rightarrow$ pick 2nd class