

- Classification bayésienne

- classification, deux classes (1, 0), une variable (X)

- probabilités **a-postérieures**:

$$p_1(x) = P(Y = 1|X = x) = \mu(x) = E[Y|X = x]$$

$$p_0(x) = P(Y = 0|X = x)$$

- décision optimale:

$$f^*(x) = \begin{cases} 1 & \text{si } p_1(x) \geq p_0(x) \\ 0 & \text{si } p_1(x) < p_0(x) \end{cases} = \begin{cases} 1 & \text{si } \mu(x) \geq \frac{1}{2} \\ 0 & \text{si } \mu(x) < \frac{1}{2} \end{cases}$$

- Le théorème de Bayes

- probabilités **a-priori**: $P(Y = C_i)$

- probabilités **conditionnelles de classe**: $p(X|Y = C_i)$

- théorème:

$$p(C_i|X) = \frac{p(X|C_i)p(C_i)}{p(X)}$$

ou

$$p(X) = \sum_{i=1}^N p(X|C_i)p(C_i)$$

- Régression quadratique

- $R(f) = E[L(Z, f)] = E[(f(X) - Y)^2]$

- $\mu(x) = E[Y|X = x]$

- fonction optimale:

$$f^*(x) = \mu(x)$$

- Théorème

- pour une fonction de décision $f: \mathbb{R} \rightarrow \{0, 1\}$ quelconque:

$$P(f^*(X) \neq Y) \leq P(f(X) \neq Y)$$

- $f^*(x)$: **décision de Bayes**

- $R^* = R(f^*) = P(f^*(X) \neq Y)$: **erreur/risque de Bayes**

- Décision de Bayes:

$$f^*(x) = C_i \text{ si } \frac{p(x|C_i)p(C_i)}{p(x)} \geq \frac{p(x|C_j)p(C_j)}{p(x)} \text{ pour tout } j = 1, \dots, N$$

ou

$$f^*(x) = C_i \text{ si } p(x|C_i)p(C_i) \geq p(x|C_j)p(C_j) \text{ pour tout } j = 1, \dots, N$$

- avantage**: $P(Y = C_i)$ et $p(X|Y = C_i)$ sont **plus faciles à estimer** que $P(Y = C_i|X = x)$ directement

- Théorème

- Pour une fonction de décision $f: \mathbb{R} \rightarrow \mathbb{R}$ quelconque:

$$E[(\mu(X) - Y)^2] \leq E[(f(X) - Y)^2]$$

- $\mu(x)$: **fonction de régression**

- dilemme biais-variance

- Fonctions discriminantes

- $g_1(\mathbf{x}), g_2(\mathbf{x}), \dots, g_N(\mathbf{x}) : \mathbb{R}^d \mapsto \mathbb{R}$

- fonction de décision:

$$f(\mathbf{x}) = C_i \text{ si } g_i(\mathbf{x}) \geq g_j(\mathbf{x}) \text{ pour tout } j = 1, \dots, N$$

- régions de décision: R_1, R_2, \dots, R_N

$$\mathbf{x} \in R_i \text{ si } f(\mathbf{x}) = C_i$$

- frontière de décision: $\mathcal{F} \subset \mathbb{R}^d$

$$\mathcal{F} = \{\mathbf{x} \mid \exists i, j : i \neq j, g_i(\mathbf{x}) = g_j(\mathbf{x})\}$$

Théorie de Bayes

- La densité normale (gaussienne)

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{x} - \mu)' \Sigma^{-1} (\mathbf{x} - \mu) \right]$$

- μ : vecteur de moyenne

- Σ : matrice de covariance (positive semidefinite)

- $|\Sigma|$: déterminante (≥ 0)

- Σ^{-1} : inverse (existe)

Théorie de Bayes

- Cas 1: $\Sigma_i = \sigma^2 \mathbf{I}$

- fonctions discriminantes linéaires: $g'_i(\mathbf{x}) = \mathbf{w}'_i \mathbf{x} + w_{i0}$

- vecteur de poids: $\mathbf{w}_i = \frac{1}{2\sigma^2} \mu_i$

- seuil ou biais: $w_{i0} = -\frac{1}{2\sigma^2} \mu'_i \mu_i + \ln P(Y = C_i)$

- frontière de décision linéaire entre R_i et R_j
 $(\mu_i - \mu_j)' (\mathbf{x} - \mathbf{x}_0)$

- $\mathbf{x}_0 = \frac{1}{2} (\mu_i + \mu_j) + \left[\frac{\sigma^2}{\|\mu_i - \mu_j\|^2} \ln \frac{P(Y = C_i)}{P(Y = C_j)} \right] (\mu_i - \mu_j)$

- $P(Y = C_i) = P(Y = C_j)$: classifieur de plus proche voisin

- Deux classes $\{-1, 1\}$

- $g(\mathbf{x}) = g_1(\mathbf{x}) - g_{-1}(\mathbf{x})$

- fonction de décision: $f(\mathbf{x}) = \text{signe}(g(\mathbf{x}))$

- régions de décision: $R_1 = \{\mathbf{x} : g(\mathbf{x}) \geq 0\}, R_{-1} = \{\mathbf{x} : g(\mathbf{x}) < 0\}$

- frontière de décision: $\mathcal{F} = \{\mathbf{x} \mid g(\mathbf{x}) = 0\}$

Théorie de Bayes

- Fonctions discriminantes

- $g_i(\mathbf{x}) = \ln p_i(\mathbf{x}) = \ln p(\mathbf{x} \mid Y = C_i) + \ln P(Y = C_i)$

- $g_i(\mathbf{x}) = -\frac{1}{2} (\mathbf{x} - \mu_i)' \Sigma_i^{-1} (\mathbf{x} - \mu_i) - \frac{d}{2} \ln 2\pi - \frac{1}{2} \ln |\Sigma_i| + \ln P(Y = C_i)$

- Cas 1: $\Sigma_i = \sigma^2 \mathbf{I}$

- $|\Sigma_i| = \sigma^{2d}$

- $\Sigma^{-1} = \frac{1}{\sigma^2} \mathbf{I}$

$$\begin{aligned} g_i(\mathbf{x}) &= -\frac{\|\mathbf{x} - \mu_i\|^2}{2\sigma^2} + \ln P(Y = C_i) + \text{const.} \\ &= -\frac{1}{2\sigma^2} (\mathbf{x}' \mathbf{x} - 2\mu'_i \mathbf{x} + \mu'_i \mu_i) + \ln P(Y = C_i) + \text{const.} \end{aligned}$$

Théorie de Bayes

- Cas 2: $\Sigma_i = \Sigma$

- $g_i(\mathbf{x}) = -\frac{1}{2} (\mathbf{x} - \mu_i)' \Sigma^{-1} (\mathbf{x} - \mu_i) + \ln P(Y = C_i) + \text{const.}$

- $g'_i(\mathbf{x}) = \mathbf{w}'_i \mathbf{x} + w_{i0}$

- vecteur de poids: $\mathbf{w}_i = \Sigma^{-1} \mu_i$

- seuil ou biais: $w_{i0} = -\frac{1}{2} \mu'_i \Sigma^{-1} \mu_i + \ln P(Y = C_i)$

• Cas 2: $\Sigma_i = \Sigma$

- **frontière de décision** entre R_i et R_j :

$$\Sigma^{-1}(\mu_i - \mu_j)'(\mathbf{x} - \mathbf{x}_0)$$

- $\mathbf{x}_0 = \frac{1}{2}(\mu_i + \mu_j) + \left[\frac{1}{(\mu_i - \mu_j)' \Sigma^{-1} (\mu_i - \mu_j)} \ln \frac{P(Y = C_i)}{P(Y = C_j)} \right] (\mu_j - \mu_i)$

- $P(Y = C_i) = P(Y = C_j)$: classifieur de plus proche voisin selon la **distance de Mahalanobis**:

$$d(\mathbf{x}, \mu) = \sqrt{(\mathbf{x} - \mu)' \Sigma^{-1} (\mathbf{x} - \mu)}$$

• Cas 3: $\Sigma_i = \text{arbitraire}$

- $g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \mu_i)' \Sigma_i^{-1} (\mathbf{x} - \mu_i) - \frac{1}{2} \ln |\Sigma_i| + \ln P(Y = C_i) + \text{const.}$

- fonctions discriminantes **quadratiques**: $g_i = \mathbf{x}' \mathbf{W}_i \mathbf{x} + \mathbf{w}_i' \mathbf{x} + w_{i0}$

- vecteur de **pooids**: $\mathbf{w}_i = \Sigma_i^{-1} \mu_i$

- $\mathbf{W}_i = -\frac{1}{2} \Sigma_i^{-1}$

- **seuil ou biais**: $w_{i0} = -\frac{1}{2} \mu_i' \Sigma_i^{-1} \mu_i - \frac{1}{2} \ln |\Sigma_i| + \ln P(Y = C_i)$

- **frontière de décision** entre R_i et R_j : **hyperquadriques**