$$e_1(\theta) = \mathbb{P}\left[G_\theta(X) < 0 \mid X \sim \mathcal{N}_d(\mu_1, \Sigma)\right]$$
$$= \Phi\left(-\frac{1}{2}D - \log\left(\frac{p_1}{p_0}\right) D^{-1}\right),$$

et $$\boxed{\begin{aligned} R_\theta(g^*) = p_0 \left(1 - \Phi\left(\frac{1}{2}D - \log\left(\frac{p_1}{p_0}\right) D^{-1}\right)\right) \\ + p_1\, \Phi\left(-\frac{1}{2}D - \log\left(\frac{p_1}{p_0}\right) D^{-1}\right) \end{aligned}}$$

4) <u>Estimation de la règle de classement</u>

On a supposé jusqu'à présent θ connu. On va à présent proposer un estimateur et on proposera le classifieur plug-in où $g^*(x) = \mathbb{1}_{G_\theta(x) > 0}$ sera remplacé par $\hat{g}(x) = \mathbb{1}_{G_{\hat{\theta}}(x) > 0}$ où $\hat{\theta}$ est l'estimateur de θ.

On estime θ par maximum de vraisemblance sur l'échantillon d'apprentissage $\mathcal{D}_n = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$.

On note f_θ la loi jointe de (X, Y) et ℓ_n la vrais. des données.

On a :

$$\ell_n(\theta) = \sum_{i=1}^n \ln f_\theta(X_i, Y_i) = \sum_{i : Y_i = 0} \log(p_0) + \log(f_0(X_i))$$
$$+ \sum_{i : Y_i = 1} \log(p_1) + \log f_1(X_i)$$

$$= \sum_{k=0}^{1} \sum_{i : Y_i = k} \Big\{ \log p_k - \frac{d}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma_k|$$
$$- \frac{1}{2}(X_i - \mu_k)^t \Sigma^{-1}(X_i - \mu_k)\Big\}$$

$$= \sum_{k=0}^{1}\left(n_k\left(\log p_k - \frac{1}{2}\log|\Sigma_k|\right)\right) - \frac{nd}{2}\log(2\pi)$$
$$- \frac{1}{2}\sum_{k=0}^{1}\sum_{i : Y_i = k}(X_i - \mu_k)^t \Sigma^{-1}(X_i - \mu_k),$$

où $n_k = \sum_{i=1}^{n} \mathbb{1}_{Y_i = k}$.

On a :

$$\frac{\partial \ln(\theta)}{\partial \mu_k} = -\frac{1}{2} \sum_{i:Y_i=k} 2\, \Sigma^{-1}(X_i - \mu_k) = -\Sigma^{-1}\Big(\sum_{i:Y_i=k} X_i - n_k \mu_k\Big)$$

d'où $\frac{\partial \ln(\theta)}{\partial \mu_k} = 0 \iff \boxed{\mu_k = \frac{1}{n_k} \sum_{i:Y_i=k} X_i}$

Pour calculer $\frac{\partial \ln(\theta)}{\partial \Sigma}$ on écrit d'abord :

$$(X_i - \mu_k)^t \Sigma^{-1} (X_i - \mu_k) = \mathrm{Tr}\Big(\underbrace{(X_i - \mu_k)^t \Sigma^{-1} (X_i - \mu_k)}\Big)$$

car $\cdot \in \mathbb{R}$

$$= \mathrm{Tr}\Big((X_i - \mu_k)(X_i - \mu_k)^t \Sigma^{-1}\Big),$$ car $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$

et on remarque que : $|\Sigma| = \frac{1}{|\Sigma^{-1}|}$, $\log|\Sigma| = -\log|\Sigma^{-1}|$

et $\frac{\partial \log|\Sigma^{-1}|}{\partial \Sigma^{-1}} = \Sigma$, $\frac{\partial}{\partial \Sigma^{-1}} \mathrm{Tr}\Big((X_i - \mu_k)(X_i - \mu_k)^t \Sigma^{-1}\Big) = (X_i - \mu_k)(X_i - \mu_k)^t$

Donc,

$$\frac{\partial \ln(\theta)}{\partial \Sigma^{-1}} = \sum_{k=0}^{1} \frac{n_k}{2} \Sigma - \frac{1}{2} \sum_{k=0}^{1} \sum_{i:Y_i=k} (X_i - \mu_k)(X_i - \mu_k)^t$$

$$= 0 \iff \boxed{\Sigma = \frac{1}{n} \sum_{k=0}^{1} \sum_{i:Y_i=k} (X_i - \mu_k)(X_i - \mu_k)^t}$$

Pour trouver p_k il faut faire la maximisation sous la contrainte $p_0 + p_1 = 1$. On écrit le Lagrangien :

$\ln(\theta, \lambda) = \ln(\theta) - \lambda(p_0 + p_1 - 1)$

$\frac{\partial \ln(\theta, \lambda)}{\partial p_k} = \frac{n_k}{p_k} - \lambda = 0 \iff \lambda = \frac{n_k}{p_k}$; $k = 0, 1$.

Donc $\lambda = \frac{n_0}{p_0} = \frac{n_1}{p_1}$, $\underbrace{p_1}_{1-p_0} n_0 = n_1 p_0$, $n_0 = (n_0+n_1)p_0$

$$p_0 = \frac{n_0}{n_0+n_1} = \boxed{\frac{n_0}{n}} \quad \text{et} \quad p_1 = 1-p_0 = \boxed{\frac{n_1}{n}}.$$

<u>Conclusion</u> : les estimateurs du max. de vrais. sont :

$$\hat{\mu_k} = \frac{1}{n_k} \sum_{i:Y_i=k} X_i \,, \quad \hat{\Sigma} = \frac{1}{n} \sum_{k=0}^{1} \sum_{i:Y_i=k} (X_i-\mu_k)(X_i-\mu_k)^t$$

$$\hat{p_k} = \frac{n_k}{n}.$$

<u>Remarque</u>: l'estimateur du max. de vrais. $\hat{\Sigma}$ est biaisé. On peut alors proposer $\tilde{\Sigma} = \frac{n}{n-2} \hat{\Sigma}$ comme estimateur ss biais.