Theoretical foundations of Machine Learning – Answers to TD

Antoine Roche

November 18, 2026

Disclaimer. This is not an official solution sheet made by your teacher Vianney Perchet, this document is only meant to help you in your studies. The solutions have to be taken with a grain of salt. If you see what you think is a typo or mistake (this is totally possible!), feel free to reach out to me at antoine.roche@ensae.fr !

TD 1

Exercise 1

1. Recall that $N_x = \sum_{i=1}^n \mathbb{1}_{\{X_i=x\}}$, that $\hat{\eta}(x) = \frac{1}{N_x}\sum_{i=1}^n Y_i \mathbb{1}_{\{X_i=x\}}$ when $N_x > 0$, and that $\hat{\eta}(x)$ takes a fixed value in $[0,1]$ when $N_x = 0$.

   Since N_x is a function of $X_1, \dots, X_n$, the indicators $\mathbb{1}_{\{N_x>0\}}$ and $\mathbb{1}_{\{N_x=0\}}$ can be taken out of the conditional expectation :

$$\begin{aligned}\mathbb{E}\left(|\hat{\eta}(x) - \eta(x)| \mid X_1, \dots, X_n\right) &= \mathbb{E}\left(|\hat{\eta}(x) - \eta(x)| \left(\mathbb{1}_{\{N_x>0\}} + \mathbb{1}_{\{N_x=0\}}\right) \mid X_1, \dots, X_n\right) \\ &= \mathbb{1}_{\{N_x>0\}}\, \mathbb{E}\left(|\hat{\eta}(x) - \eta(x)| \mid X_1, \dots, X_n\right) \\ &\quad + \mathbb{1}_{\{N_x=0\}}\, \mathbb{E}\left(|\hat{\eta}(x) - \eta(x)| \mid X_1, \dots, X_n\right)\end{aligned}$$

   Term on $\{N_x = 0\}$. Both $\hat{\eta}(x)$ and $\eta(x)$ lie in $[0,1]$, so $|\hat{\eta}(x) - \eta(x)| \leqslant 1$ and

$$\mathbb{1}_{\{N_x=0\}}\, \mathbb{E}\left(|\hat{\eta}(x) - \eta(x)| \mid X_1, \dots, X_n\right) \leqslant \mathbb{1}_{\{N_x=0\}}$$

   Term on $\{N_x > 0\}$. Since the pairs (X_i, Y_i) are i.i.d., conditionally on $X_1, \dots, X_n$ the Y_i are independent, with $\mathbb{E}(Y_i \mid X_1, \dots, X_n) = \eta(X_i)$ and $\mathrm{Var}(Y_i \mid X_1, \dots, X_n) = \eta(X_i)(1 - \eta(X_i))$. Hence, on $\{N_x > 0\}$,

$$\begin{aligned}\mathbb{E}\left(\hat{\eta}(x) \mid X_1, \dots, X_n\right) &= \frac{1}{N_x}\sum_{i=1}^n \mathbb{1}_{\{X_i=x\}}\eta(X_i) = \frac{1}{N_x}\sum_{i=1}^n \mathbb{1}_{\{X_i=x\}}\eta(x) = \eta(x) \\ \mathrm{Var}\left(\hat{\eta}(x) \mid X_1, \dots, X_n\right) &= \frac{1}{N_x^2}\sum_{i=1}^n \mathbb{1}_{\{X_i=x\}}\eta(x)(1-\eta(x)) = \frac{\eta(x)(1-\eta(x))}{N_x} \leqslant \frac{1}{4N_x}\end{aligned}$$

   By Cauchy-Schwarz (or Jensen), we deduce

$$\begin{aligned}\mathbb{1}_{\{N_x>0\}}\, \mathbb{E}\left(|\hat{\eta}(x) - \eta(x)| \mid X_1, \dots, X_n\right) &\leqslant \mathbb{1}_{\{N_x>0\}} \left(\mathbb{E}\left((\hat{\eta}(x) - \eta(x))^2 \mid X_1, \dots, X_n\right)\right)^{\frac{1}{2}} \\ &= \mathbb{1}_{\{N_x>0\}} \left(\mathrm{Var}\left(\hat{\eta}(x) \mid X_1, \dots, X_n\right)\right)^{\frac{1}{2}} \\ &\leqslant \frac{1}{2\sqrt{N_x}}\mathbb{1}_{\{N_x>0\}} \leqslant \frac{1}{\sqrt{N_x}}\mathbb{1}_{\{N_x>0\}}\end{aligned}$$

   Conclusion. Combining both terms,

$$\mathbb{E}\left(|\hat{\eta}(x) - \eta(x)| \mid X_1, \dots, X_n\right) \leqslant \frac{1}{\sqrt{N_x}}\mathbb{1}_{\{N_x>0\}} + \mathbb{1}_{\{N_x=0\}}$$

Denote $\mathcal{D}_n = ((X_1, Y_1), \dots, (X_n, Y_n))$ and let (X, Y) be a new pair, independent of $\mathcal{D}_n$, with the same distribution. With $\hat{f}(x) = \mathbb{1}_{\{\hat{\eta}(x) \geqslant 1/2\}}$ and $f^*(x) = \mathbb{1}_{\{\eta(x) \geqslant 1/2\}}$, we have $R(\hat{f}) = \mathbb{P}(Y \neq \hat{f}(X) \mid \mathcal{D}_n)$ and $R^* = \mathbb{P}(Y \neq f^*(X))$.

Pointwise excess risk. Since (X, Y) is independent of $\mathcal{D}_n$, conditionally on $(X, \mathcal{D}_n)$ the variable Y is Bernoulli with parameter $\eta(X)$. Hence

$$\mathbb{P}(Y \neq \hat{f}(X) \mid X, \mathcal{D}_n) = \eta(X)\mathbb{1}_{\{\hat{f}(X)=0\}} + (1 - \eta(X))\mathbb{1}_{\{\hat{f}(X)=1\}},$$

and similarly for f^*. The difference of the two expressions is zero when $\hat{f}(X) = f^*(X)$. Otherwise it equals $2\eta(X) - 1 \geqslant 0$ if $f^*(X) = 1$, and $1 - 2\eta(X) > 0$ if $f^*(X) = 0$. Therefore

$$\mathbb{P}(Y \neq \hat{f}(X) \mid X, \mathcal{D}_n) - \mathbb{P}(Y \neq f^*(X) \mid X) = |2\eta(X) - 1|\mathbb{1}_{\{\hat{f}(X) \neq f^*(X)\}}.$$

Link with $\hat{\eta}$. If $\hat{f}(X) \neq f^*(X)$, then $\hat{\eta}(X)$ and $\eta(X)$ lie on either side of $1/2$, so $|2\eta(X) - 1| = 2|\eta(X) - \frac{1}{2}| \leqslant 2|\hat{\eta}(X) - \eta(X)|$. Hence

$$|2\eta(X) - 1|\mathbb{1}_{\{\hat{f}(X) \neq f^*(X)\}} \leqslant 2|\hat{\eta}(X) - \eta(X)|.$$

Conclusion. Taking expectations, splitting according to whether $N_X > 0$, applying question 1 at the point $x = X$ on $\{N_X > 0\}$ (which is legitimate since X is independent of $\mathcal{D}_n$), and using $|2\eta(X) - 1| \leqslant 1$ on $\{N_X = 0\}$,

$$\begin{aligned}
\mathbb{E}\left[R(\hat{f}) - R^*\right] &= \mathbb{E}\left[|2\eta(X) - 1|\mathbb{1}_{\{\hat{f}(X) \neq f^*(X)\}}\right] \\
&\leqslant 2\mathbb{E}\left[\mathbb{1}_{\{N_X > 0\}}\mathbb{E}\left(|\hat{\eta}(X) - \eta(X)| \mid X, X_1, \dots, X_n\right)\right] + \mathbb{P}(N_X = 0) \\
&\leqslant 2\mathbb{E}\left[\frac{1}{\sqrt{N_X}}\mathbb{1}_{\{N_X > 0\}}\right] + \mathbb{P}(N_X = 0)
\end{aligned}$$

2. By the law of large numbers, for every x such that $\mathbb{P}(X_1 = x) > 0$ we have $N_x \to \infty$ a.s. Hence $\mathbb{P}(N_x = 0) \to 0$ and moreover, by the dominated convergence theorem,

$$\mathbb{E}\left[\frac{1}{\sqrt{N_x}}\mathbb{1}_{\{N_x > 0\}}\right] \to 0 \text{ since } 0 \leqslant \frac{1}{\sqrt{N_x}}\mathbb{1}_{\{N_x > 0\}} \leqslant 1$$

Since X is independent of $\mathcal{D}_n$ and almost surely takes a value x such that $\mathbb{P}(X_1 = x) > 0$, we deduce

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{\sqrt{N_X}}\mathbb{1}_{\{N_X > 0\}}\right] &= \sum_{x : \mathbb{P}(X_1 = x) > 0} \mathbb{P}(X = x)\,\mathbb{E}\left[\frac{1}{\sqrt{N_x}}\mathbb{1}_{\{N_x > 0\}}\right] \to 0 \\
\mathbb{P}(N_X = 0) &= \sum_{x : \mathbb{P}(X_1 = x) > 0} \mathbb{P}(X = x)\,\mathbb{P}(N_x = 0) \to 0
\end{aligned}$$

as finite sums of sequences converging to 0.

3. By question 2, whatever the distribution of the data, the risk of our estimator converges to the Bayes risk, so $\hat{f}$ is universally consistent.

Exercise 2

We expand the expression of the expectation.[1]

$$\begin{aligned}
\mathbb{E}_{\theta \sim \pi(\cdot|X)}[L(\theta, a)] &= \mathbb{E}_{\theta \sim \pi(\cdot|X)}\left[k_1(a - \theta)\mathbb{1}_{\{\theta \leqslant a\}} + k_2(\theta - a)\mathbb{1}_{\{\theta > a\}}\right] \\
&= k_1\mathbb{E}_{\theta \sim \pi(\cdot|X)}\left[(a - \theta)\mathbb{1}_{\{\theta \leqslant a\}}\right] + k_2\mathbb{E}_{\theta \sim \pi(\cdot|X)}\left[(\theta - a)\mathbb{1}_{\{\theta > a\}}\right]
\end{aligned}$$

[1] **Reminder from Stat 1.** We assume that the random variables X and θ have a density with respect to the Lebesgue measure. The (Bayesian) setting of the exercise is the following :

$$\begin{aligned}
\theta &\sim \pi \\
X \mid \theta &\sim \mathbb{P}_\theta
\end{aligned}$$

We use the following result : for a nonnegative random variable Z, $\mathbb{E}[Z] = \int_0^{+\infty} \mathbb{P}(Z \geqslant t)\mathrm{d}t$. We obtain

$$\begin{aligned}
\mathbb{E}_{\theta\sim\pi(\cdot|X)}[L(\theta,a)] &= k_1 \int_0^{+\infty} \mathbb{P}_{\theta\sim\pi(\cdot|X)} \left(\mathbb{1}_{\{\theta\leqslant a\}}(a-\theta) \geqslant t\right) \mathrm{d}t \\
&\quad + k_2 \int_0^{+\infty} \mathbb{P}_{\theta\sim\pi(\cdot|X)} \left(\mathbb{1}_{\{\theta> a\}}(\theta-a) \geqslant t\right) \mathrm{d}t \\
&= k_1 \int_0^{+\infty} \mathbb{P}_{\theta\sim\pi(\cdot|X)}(\theta \leqslant a-t)\mathrm{d}t \\
&\quad + k_2 \int_0^{+\infty} \mathbb{P}_{\theta\sim\pi(\cdot|X)}(\theta \geqslant a+t)\mathrm{d}t \\
&= k_1 \int_{-\infty}^{a} \mathbb{P}_{\theta\sim\pi(\cdot|X)}(\theta \leqslant u)\mathrm{d}u \\
&\quad + k_2 \int_a^{+\infty} \mathbb{P}_{\theta\sim\pi(\cdot|X)}(\theta > u)\mathrm{d}u
\end{aligned}$$

This function is differentiable almost everywhere, and at its points of differentiability, its derivative equals

$$k_1 \mathbb{P}_{\theta\sim\pi(\cdot|X)}(\theta \leqslant a) - k_2 \mathbb{P}_{\theta\sim\pi(\cdot|X)}(\theta > a) = -(k_1+k_2)\mathbb{P}_{\theta\sim\pi(\cdot|X)}(\theta > a) + k_1.$$

We note that its derivative is negative for every a such that $\mathbb{P}_{\theta\sim\pi(\cdot|X)}(\theta > a) > k_1/(k_1+k_2)$, and that it is positive for every a such that $\mathbb{P}_{\theta\sim\pi(\cdot|X)}(\theta > a) < k_1/(k_1+k_2)$.
The risk is minimal at $\hat{\theta}$ such that $\mathbb{P}_{\theta\sim\pi(\cdot|X)}(\theta > \hat{\theta}) = k_1/(k_1+k_2)$, that is $\mathbb{P}_{\theta\sim\pi(\cdot|X)}(\theta \leqslant \hat{\theta}) = k_2/(k_1+k_2)$. The Bayes estimator is therefore the quantile of level $k_2/(k_1+k_2)$ of the posterior distribution $\pi(\cdot \mid X)$.

Exercise 3

1. By Bayes' formula for conditional probability densities,

$$\begin{aligned}
\eta(x) &= \mathbb{P}(Y=1 \mid X=x) \\
&= \frac{f_{Y=1}(x)\mathbb{P}(Y=1)}{f_{Y=1}(x)\mathbb{P}(Y=1) + f_{Y=0}(x)\mathbb{P}(Y=0)} \\
&= \frac{f_1(x)p}{f_1(x)p + f_0(x)(1-p)}
\end{aligned}$$

By abuse of notation, we denote by $\pi(\cdot \mid x)$ the posterior distribution of θ, that is the distribution of θ conditionally on $X = x$. Recall that this distribution is obtained using Bayes' formula. In the Stat1 course, denoting by $\mathcal{L}$ the likelihood, the posterior distribution is given by

$$\pi(\mathrm{d}\theta \mid x) = \frac{\mathcal{L}(\theta;x)\pi(\mathrm{d}\theta)}{\int \mathcal{L}(\theta';x)\pi(\mathrm{d}\theta')}$$

You can easily convince yourself that with the notation of the problem statement, we obtain the following formula :

$$\pi(\mathrm{d}\theta \mid x) = \frac{\mathbb{P}_\theta(x)\pi(\mathrm{d}\theta)}{\int \mathbb{P}_{\theta'}(x)\pi(\mathrm{d}\theta')}$$

The Bayes estimator minimizes the risk associated with the loss L under the posterior distribution. Be careful, L denotes here the loss and not the likelihood! Conditionally on X, the Bayes estimator satisfies

$$\begin{aligned}
\hat{\theta} &= \arg\min_a R(a) \\
&= \arg\min_a \mathbb{E}_{\theta\sim\pi(\cdot|X)}[L(\theta,a)]
\end{aligned}$$

2. We showed in the course that the optimal classifier, also called the Bayes predictor, is the plug-in estimator obtained from the conditional expectation of Y given X.

   It is therefore $f^*(x) = \mathbb{1}_{\left\{\eta(x) > \frac{1}{2}\right\}} = \mathbb{1}_{\{pf_1(x) > (1-p)f_0(x)\}}$.

3. Recall that the Bayes risk is $L^* = \mathbb{P}(Y \neq f^*(X))$.

   (a) Let us compute $f^*(x)$ for the different values of x :
   - For $x \in (\theta_0, \theta_1]$, $f_0(x) = 0 < f_1(x)$ so $f^*(x) = 1$.
   - For $x \in [0, \theta_0]$, $f^*(x) = \mathbb{1}_{\left\{\frac{p}{\theta_1} > \frac{1-p}{\theta_0}\right\}}$. If $\theta_0/\theta_1 > (1-p)/p$, $f^*(x) = 1$, otherwise $f^*(x) = 0$.
   - For $x \notin [0, \theta_1]$, $f_0(x) = f_1(x) = 0$ so $f^*(x) = 0$.

   The Bayes predictor is therefore given by the following formula :
   - If $\theta_0/\theta_1 > (1-p)/p$, $f^*(x) = \mathbb{1}_{[0,\theta_1]}(x)$;
   - If $\theta_0/\theta_1 \leqslant (1-p)/p$, $f^*(x) = \mathbb{1}_{(\theta_0,\theta_1]}(x)$.

   We then compute the Bayes risk L^* according to the relative values of the ratios θ_0/θ_1 and $(1-p)/p$.
   Case 1 : $\theta_0/\theta_1 > (1-p)/p$. Then we showed that $f^*(x) = 1$ for every $x \in [0, \theta_1]$, and $X \in [0, \theta_1]$ almost surely, so

$$\begin{aligned} L^* &= \mathbb{P}[Y \neq f^*(X)] \\ &= \mathbb{P}[Y = 0] \\ &= 1 - p \end{aligned}$$

   Case 2 : $\theta_0/\theta_1 \leqslant (1-p)/p$. Then we showed that $f^*(x) = \mathbb{1}_{(\theta_0,\theta_1]}(x)$, so

$$\begin{aligned} L^* &= \mathbb{P}(Y = 1 \text{ and } f^*(X) = 0) + \mathbb{P}(Y = 0 \text{ and } f^*(X) = 1) \\ &= \mathbb{P}(f^*(X) = 0 \mid Y = 1)\mathbb{P}(Y = 1) + \mathbb{P}(f^*(X) = 1 \mid Y = 0)\mathbb{P}(Y = 0) \\ &= \mathbb{P}_{X\sim f_1}[X \leqslant \theta_0] \cdot p + (1-p)\mathbb{P}_{X\sim f_0}[X > \theta_0] \\ &= p\frac{\theta_0}{\theta_1} \end{aligned}$$

   We therefore have $L^* = p \min\left(\frac{\theta_0}{\theta_1}, \frac{1-p}{p}\right)$.

   (b) Let us compute $f^*(x)$ for the different values of x :
   - For $x \in (\theta_0, \theta_1]$, $f_0(x) = 0 < f_1(x)$ so $f^*(x) = 1$.
   - For $x \in [0, \theta_0)$, $f_1(x) = 0$ so $f^*(x) = 0$.
   - For $x \notin [0, \theta_1]$, $f_0(x) = f_1(x) = 0$ so $f^*(x) = 0$.

   We therefore have $f^*(x) = \mathbb{1}_{\{\theta_0 < x \leqslant \theta_1\}}$ for every $x \neq \theta_0$. Let us now compute the Bayes risk.

$$\begin{aligned} L^* &= p\mathbb{P}_{X\sim f_1}[X \leqslant \theta_0] + (1-p)\mathbb{P}_{X\sim f_0}[X > \theta_0] \\ &= 0. \end{aligned}$$

   (c) Let us compute the Bayes predictor. For $x \geqslant 0$, we have

$$\begin{aligned} f^*(x) &= \mathbb{1}_{\{pf_1(x) > (1-p)f_0(x)\}} \\ &= \mathbb{1}_{\left\{p\theta_1 e^{-\theta_1 x} > (1-p)\theta_0 e^{-\theta_0 x}\right\}} \\ &= \mathbb{1}_{\left\{\frac{p\theta_1}{(1-p)\theta_0} > e^{(\theta_1-\theta_0)x}\right\}} \\ &= \mathbb{1}_{\left\{x < \frac{1}{\theta_1-\theta_0}\log\left(\frac{p\theta_1}{(1-p)\theta_0}\right)\right\}} \end{aligned}$$

We note that if $\frac{p\theta_1}{(1-p)\theta_0} \leqslant 1$, that is if $p/(1-p) \leqslant \theta_0/\theta_1$, then $f^*(x) = 0$ for every $x > 0$.

Let us now compute the Bayes risk. To simplify notation, we set

$$x_0 = \left(\frac{1}{\theta_1 - \theta_0} \log \left(\frac{p\theta_1}{(1-p)\theta_0} \right) \right)_+$$

We therefore have $f^*(x) = \mathbb{1}_{\{x<x_0\}}$, and

$$\begin{aligned} L^* &= p\mathbb{P}_{X\sim f_1}[X \geqslant x_0] + (1-p)\mathbb{P}_{X\sim f_0}[X < x_0] \\ &= pe^{-\theta_1 x_0} + (1-p)\left(1 - e^{-\theta_0 x_0}\right) \end{aligned}$$

Case 1 : If $p/(1-p) \leqslant \theta_0/\theta_1$, then $x_0 = 0$ and

$$L^* = p$$

Case 2 : If $p/(1-p) > \theta_0/\theta_1$, then[2]

$$x_0 = \frac{1}{\theta_1 - \theta_0} \log \left(\frac{p\theta_1}{(1-p)\theta_0} \right)$$

Thus,

$$L^* = p\left(\frac{p\theta_1}{(1-p)\theta_0} \right)^{-\frac{\theta_1}{\theta_1-\theta_0}} + (1-p)\left(1 - \left(\frac{p\theta_1}{(1-p)\theta_0} \right)^{-\frac{\theta_0}{\theta_1-\theta_0}} \right)$$

(d) Let us compute the Bayes predictor. For $x \in \mathbb{R}$, we have

$$\begin{aligned} f^*(x) &= \mathbb{1}_{\{pf_1(x) > (1-p)f_0(x)\}} \\ &= \mathbb{1}_{\left\{ p\frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\theta_1)^2}{2\sigma_1^2}} > (1-p)\frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{(x-\theta_0)^2}{2\sigma_0^2}} \right\}} \\ &= \mathbb{1}_{\left\{ \frac{p}{1-p}\cdot\frac{\sigma_0}{\sigma_1} > e^{\frac{(x-\theta_1)^2}{2\sigma_1^2} - \frac{(x-\theta_0)^2}{2\sigma_0^2}} \right\}} \end{aligned}$$

For the Bayes risk, we have :

$$\begin{aligned} L^* &= \mathbb{E}[\min(\eta(X), 1-\eta(X))] \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \min\left(\frac{p}{\sigma_1} e^{-\frac{(x-\theta_1)^2}{2\sigma_1^2}}, \frac{1-p}{\sigma_0} e^{-\frac{(x-\theta_0)^2}{2\sigma_0^2}} \right) \mathrm{d}x \end{aligned}$$

Case 1 : The following equation has no solution :

$$\frac{(x-\theta_1)^2}{2\sigma_1^2} - \frac{(x-\theta_0)^2}{2\sigma_0^2} = \ln\left(\frac{p}{1-p} \cdot \frac{\sigma_0}{\sigma_1} \right)$$

We note that at $+\infty$ and $-\infty$:

$$\frac{(x-\theta_1)^2}{2\sigma_1^2} - \frac{(x-\theta_0)^2}{2\sigma_0^2} \sim \frac{x^2}{2}\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2} \right)$$

[2]Recall that the cumulative distribution function of an exponential distribution with parameter λ is given by $F(t) = 1 - e^{-\lambda t}$.

so if $\sigma_1 > \sigma_0$, $\forall x$,

$$\frac{p}{1-p} \cdot \frac{\sigma_0}{\sigma_1} > e^{\frac{(x-\theta_1)^2}{2\sigma_1^2} - \frac{(x-\theta_0)^2}{2\sigma_0^2}},$$

and conversely in the case where $\sigma_1 < \sigma_0$, which gives us :

$$\begin{aligned} f^*(x) &= \mathbb{1}_{\{\sigma_1 > \sigma_0\}} \\ L^* &= (1-p) \cdot \mathbb{1}_{\{\sigma_1 > \sigma_0\}} + p \cdot \mathbb{1}_{\{\sigma_0 > \sigma_1\}} \end{aligned}$$

The same conclusion holds if $\sigma_0 \neq \sigma_1$ and the equation has a single (double) root, since this only changes f^* at one point.

Case 2 : The following equation has two solutions :

$$\frac{(x-\theta_1)^2}{2\sigma_1^2} - \frac{(x-\theta_0)^2}{2\sigma_0^2} = \ln\left(\frac{p}{1-p} \cdot \frac{\sigma_0}{\sigma_1}\right)$$

which we denote by $a_1 < a_2$. Assume $\sigma_1 > \sigma_0$.

$$\begin{aligned} f^*(x) &= 1 - \mathbb{1}_{\{a_1 < x < a_2\}} \\ L^* &= \frac{1}{\sqrt{2\pi}} \left(\int_{-\infty}^{a_1} \frac{1-p}{\sigma_0} e^{-\frac{(x-\theta_0)^2}{2\sigma_0^2}} \mathrm{d}x \right. \\ &\quad + \int_{a_2}^{+\infty} \frac{1-p}{\sigma_0} e^{-\frac{(x-\theta_0)^2}{2\sigma_0^2}} \mathrm{d}x \\ &\quad \left. + \int_{a_1}^{a_2} \frac{p}{\sigma_1} e^{-\frac{(x-\theta_1)^2}{2\sigma_1^2}} \mathrm{d}x \right) \\ &= (1-p) \left(\Phi\left(\frac{a_1 - \theta_0}{\sigma_0}\right) + 1 - \Phi\left(\frac{a_2 - \theta_0}{\sigma_0}\right) \right) \\ &\quad + p \left(\Phi\left(\frac{a_2 - \theta_1}{\sigma_1}\right) - \Phi\left(\frac{a_1 - \theta_1}{\sigma_1}\right) \right) \end{aligned}$$

where Φ is the cumulative distribution function of the standard Gaussian.
If $\sigma_1 < \sigma_0$, the roles are exchanged : $f^*(x) = \mathbb{1}_{\{a_1 < x < a_2\}}$ and

$$\begin{aligned} L^* &= (1-p) \left(\Phi\left(\frac{a_2 - \theta_0}{\sigma_0}\right) - \Phi\left(\frac{a_1 - \theta_0}{\sigma_0}\right) \right) \\ &\quad + p \left(\Phi\left(\frac{a_1 - \theta_1}{\sigma_1}\right) + 1 - \Phi\left(\frac{a_2 - \theta_1}{\sigma_1}\right) \right) \end{aligned}$$

Case 3 : $\sigma_0 = \sigma_1 = \sigma$. The equation above is then affine in x and has exactly one solution

$$a = \frac{\theta_0 + \theta_1}{2} + \frac{\sigma^2}{\theta_1 - \theta_0} \ln\left(\frac{1-p}{p}\right),$$

and since $\theta_1 > \theta_0$,

$$\begin{aligned} f^*(x) &= \mathbb{1}_{\{x > a\}} \\ L^* &= p\,\Phi\left(\frac{a - \theta_1}{\sigma}\right) + (1-p) \left(1 - \Phi\left(\frac{a - \theta_0}{\sigma}\right)\right) \end{aligned}$$