# SML ASSIGNMENT 2

Détagnon Bernic GBAGUIDI

January 21, 2024

## Exercice 1 Nearest Neighbors as a Kernel Method

1. Let explain clearly why and how $\hat{f}_{kNN}(x)$ does indeed compute the predicted label of x

   $$\hat{f}_{kNN}(x) = \text{sign}(\sum_{i=1}^{n} y_i \alpha_i \mathcal{K}(x, x_i))$$

   $\mathcal{K}(x, x_i)$ is the weight attributed to the observation $x_i$ .

   It is in a such way that if $d(x, x_i) \leq d(x, x_j)$ then $\mathcal{K}(x, x_j) \leq \mathcal{K}(x, x_i)$. This mean that more close are the observation to $x$ and more bigger is it weight.

   $\sum_{i=1}^{n} y_i \alpha_i \mathcal{K}(x, x_i)$ is the summation of all the response of the observations that are in the neighborhood of $x$ multiply by their weight.

   With this principle $x$ is classify in the class of obsevations that are more close of it.

2. Let explain succinctly what$\hat{\pi}(x)$ is estimating in this case

   $$\hat{\pi}(x) = \sum_{i=1}^{n} \mathbb{1}(y_i = 1) \alpha_i \mathcal{K}(x, x_i))$$

   $\hat{\pi}(x)$ compute the sum of the weight of all observations in the neighborhood of $x$ that have response 1.

   It is an estimation of $Pr(Y = 1|x)$

3. Let explain clearly the relationship between $\hat{f}_{kNN}(x)$ and

   $$\hat{g}_{kNN}(x) = 2\mathbb{1}(\hat{\pi}(x) > \frac{1}{2}) - 1$$

- if $\hat{\pi}(x) > \frac{1}{2}$
  if $\hat{\pi}(x) > \frac{1}{2}$, we have $\hat{g}_{kNN}(x) = 2 \times 1 - 1 = 1$
  And at the same time by the way $\hat{\pi}(x)$ is defined, that's mean that ,

$$\sum_{i=1}^{n} \mathbb{1}(y_i = 1)\alpha_i \mathcal{K}(x, x_i)) > \sum_{i=1}^{n} \mathbb{1}(y_i = -1)\alpha_i \mathcal{K}(x, x_i))$$

  then, $\text{sign}(\sum_{i=1}^{n} y_i \alpha_i \mathcal{K}(x, x_i)) = 1$ i.e. $\hat{f}_{kNN}(x) =$**1**
  So if $\hat{\pi}(x) > \frac{1}{2}$ we have $\hat{g}_{kNN}(x) = \hat{f}_{kNN}(x)$
- if $\hat{\pi}(x) \leq \frac{1}{2}$
  if $\hat{\pi}(x) \leq \frac{1}{2}$ , then $\hat{g}_{kNN}(x) = 2 \times 0 - 1 = -1$
  And at the same time by the way $\hat{\pi}(x)$ is defined, that's mean that ,

$$\sum_{i=1}^{n} \mathbb{1}(y_i = 1)\alpha_i \mathcal{K}(x, x_i)) < \sum_{i=1}^{n} \mathbb{1}(y_i = -1)\alpha_i \mathcal{K}(x, x_i))$$

  then, $\text{sign}(\sum_{i=1}^{n} y_i \alpha_i \mathcal{K}(x, x_i)) = -1$ i.e. $\hat{f}_{kNN}(x) =$**-1**
  So if $\hat{\pi}(x) \leq \frac{1}{2}$ we have $\hat{g}_{kNN}(x) = \hat{f}_{kNN}(x)$
  **Conclusion :** $\hat{g}_{kNN}(x) = \hat{f}_{kNN}(x)$

4. In this case we have $\mathcal{Y} = \{0, 1\}$ because $\mathbb{1}(\hat{\pi}(x) > \frac{1}{2}) \in \{0, 1\}$

5. We can state the logistic regression function space defined as follow: Let denote

$$\hat{\pi}(x) = \frac{1}{1 + e^{-\sum_i \beta_i x_i}}$$

   We have,

$$\mathcal{H} := \left\{ f s.t, \forall x \in \mathcal{X}, f(x) = \mathbb{1}\left(\hat{\pi}(x) > \frac{1}{2}\right) \right\}$$

# Exercice 2:When can we compute the Bayes' Risk?

1. Let sketch the contour of the two classes in the plane.

   **(SEE THE RMARKDOWN CODE)**

2. Let find and write down the simplified expression of the Bayes' classifier for this task. We have,

$$
\begin{aligned}
f^*(x) =& \mathbb{1}\left(\frac{\mathbb{P}[Y=1|x]}{\mathbb{P}[Y=0|x]} \geq 1\right) \\
=& \mathbb{1}\left(ln(\mathbb{P}[Y=1|x]) - ln(\mathbb{P}[Y=0|x]) > 0\right) \\
=& \mathbb{1}\left(ln(\frac{\mathbb{P}[x|Y=1]\mathbb{P}[Y=1]}{\mathbb{P}[x]}) - ln(\frac{\mathbb{P}[x|Y=0]\mathbb{P}[Y=0]}{\mathbb{P}[x]}) > 0\right) \\
=& \mathbb{1}\left(ln(\mathbb{P}[x|Y=1]) + ln(\mathbb{P}[Y=1]) - ln(\mathbb{P}[x|Y=0]) - ln(\mathbb{P}[Y=0]) > 0\right) \\
=& \mathbb{1}\left(\frac{-1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1) + \frac{1}{2}(x-\mu_0)^T\Sigma^{-1}(x-\mu_0) + ln(\psi) - ln(1-\psi) > 0\right) \\
=& \mathbb{1}(\frac{1}{2}(-x^T\Sigma^{-1}x + x^T\Sigma^{-1}\mu_1 + \mu_1^T\Sigma^{-1}x - \mu_1^T\Sigma^{-1}\mu_1 + \\
& \qquad x^T\Sigma^{-1}x - x^T\Sigma^{-1}\mu_0 - \mu_0^T\Sigma^{-1}x + \mu_0^T\Sigma^{-1}\mu_0) + ln(\psi) - ln(1-\psi) > 0) \\
=& \mathbb{1}(\frac{1}{2}(2x^T\Sigma^{-1}\mu_1 - \mu_1^T\Sigma^{-1}\mu_1 - 2x^T\Sigma^{-1}\mu_0 + \mu_0^T\Sigma^{-1}\mu_0) + ln(\psi) - ln(1-\psi) > 0) \\
=& \mathbb{1}\left((\mu_1-\mu_0)^T\Sigma^{-1}x - \frac{1}{2}(\mu_1+\mu_0)^T\Sigma^{-1}(\mu_1-\mu_0) + ln(\psi) - ln(1-\psi) > 0\right)
\end{aligned}
$$

We can then conclude that,

$$f^*(x) = \mathbb{1}\left(\omega^T x + b > 0\right)$$

where $\omega = \Sigma^{-1}(\mu_1 - \mu_0)$ and
$b = -\frac{1}{2}(\mu_1+\mu_0)^T\Sigma^{-1}(\mu_1-\mu_0) + ln(\psi) - ln(1-\psi)$

3. Let denote $DB$ the decision boundary. We have ,

$$DB = \{x \in \mathcal{X} \quad s.t \quad \omega^T x + b = 0\}$$

where $\omega = \Sigma^{-1}(\mu_1 - \mu_0)$ and
$b = -\frac{1}{2}(\mu_1+\mu_0)^T\Sigma^{-1}(\mu_1-\mu_0) + ln(\psi) - ln(1-\psi)$
Let develop more. $\mu_0 = (-2 - 1)^T$ , $\mu_1 = (0 \quad 1)^T$ ,
we have $\mu_1 - \mu_0 = (2 \quad 2)$ and $\mu_1 + \mu_0 = (-2 \quad 0)$

We have,

$$\Sigma = \begin{pmatrix} 1 & \frac{-3}{4} \\ \frac{-3}{4} & 2 \end{pmatrix} \qquad \text{and} \qquad det(\Sigma) = \frac{23}{16} \quad \text{so,}$$

$$\Sigma^{-1} = \frac{16}{23}\begin{pmatrix} 2 & \frac{3}{4} \\ \frac{3}{4} & 1 \end{pmatrix}$$

Another side we have,

$$\begin{aligned}(\mu_1 - \mu_0)^T \Sigma^{-1} =& (2 \quad 2)\frac{16}{23}\begin{pmatrix} 2 & \frac{3}{4} \\ \frac{3}{4} & 1 \end{pmatrix} \\ =& \frac{8}{23}(11 \quad 7)\end{aligned}$$

and also,

$$\begin{aligned}(\mu_1 + \mu_0)^T \Sigma^{-1} =& (-2 \quad 0)\frac{16}{23}\begin{pmatrix} 2 & \frac{3}{4} \\ \frac{3}{4} & 1 \end{pmatrix} \\ =& \frac{8}{23}(-8 \quad -3)\end{aligned}$$

so,

$$\begin{aligned}\frac{-1}{2}(\mu_1 + \mu_0)^T \Sigma^{-1}(\mu_1 - \mu_0) =& \frac{-1}{2}\frac{8}{23}(-8 \quad -3)\begin{pmatrix} 2 \\ 2 \end{pmatrix} \\ =& \frac{88}{23}\end{aligned}$$

We can then write,

$$\omega^T x + b = \frac{88}{23}x_1 + \frac{56}{23}x_2 + \frac{88}{23} + ln(\psi) - ln(1-\psi)$$

Finally,

$$DB = \{x = (x_1, x_2) \in \mathcal{X} \quad s.t \quad \frac{88}{23}x_1 + \frac{56}{23}x_2 + \frac{88}{23} + ln(\psi) - ln(1-\psi) = 0\}$$

4. Theoretical expression of the Bayes' risk. We have,

$$R^* = \min_f \mathbb{P}[Y \neq f(X)]$$

but because we know $f^*$, we have

$$R^* = \mathbb{P}[Y \neq \mathbb{1}\left(\omega^T x + b > 0\right)]$$

$$\begin{aligned}
R^* =& \mathbb{P}[Y \neq \mathbb{1}\left(\omega^T x + b > 0\right)] \\
=& \mathbb{P}[Y = 1, \omega^T x + b \leq 0] + \mathbb{P}[Y = 0, \omega^T x + b > 0] \\
=& \mathbb{P}[\omega^T x + b \leq 0 | Y = 1]\mathbb{P}[Y = 1] + \mathbb{P}[\omega^T x + b > 0 | Y = 0]\mathbb{P}[Y = 0] \\
=& \psi\mathbb{P}[\omega^T x + b \leq 0 | Y = 1] + (1 - \psi)\mathbb{P}[\omega^T x + b > 0 | Y = 0] \\
=& \psi\mathbb{P}[\omega^T x \leq -b | Y = 1] + (1 - \psi)\mathbb{P}[\omega^T x > -b | Y = 0]
\end{aligned}$$

But $X|Y$ is bivariate gaussian distribution for which we now the parameters, therefor $\omega^T X|Y$ is also a gaussian distribution . We are going to determine the parameters.

$$\begin{aligned}
\mathrm{E}_0(\omega^T X | Y = 0) =& \omega^T \mathrm{E}_0(X | Y = 0) \\
=& \omega^T(\mu_0) \\
=& \frac{8}{23}(11 \quad 7)\begin{pmatrix} -2 \\ -1 \end{pmatrix} \\
=& \frac{-232}{23}
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}_1(\omega^T X | Y = 1) =& \omega^T \mathrm{E}_1(X | Y = 1) \\
=& \omega^T(\mu_1) \\
=& \frac{8}{23}(11 \quad 7)\begin{pmatrix} 0 \\ 1 \end{pmatrix} \\
=& \frac{56}{23}
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}(\omega^T X | Y = 1) =& \omega^T \Sigma \omega \\
=& \frac{8}{23}(11 \quad 7)\begin{pmatrix} 1 & \frac{-3}{4} \\ \frac{-3}{4} & 2 \end{pmatrix}\begin{pmatrix} 11 \\ 7 \end{pmatrix}\frac{8}{23} \\
=& \frac{288}{23}
\end{aligned}$$

With this we can deduce that,

$Z_0 = \dfrac{(\omega^T X|Y=0) - (\dfrac{-232}{23})}{\sqrt{\dfrac{288}{23}}}$ follow a standard normal distribution

and $Z_1 = \dfrac{(\omega^T X|Y=1) - \dfrac{56}{23}}{\sqrt{\dfrac{288}{23}}}$ also .

Let call $Z$ a standard normal distribution and $\phi(x)$ its CDF

We have

$$R^* = \psi\phi(\frac{-b - \dfrac{56}{23}}{\sqrt{\dfrac{288}{23}}}) + (1-\psi)(1 - \phi(\frac{-b + (\dfrac{232}{23})}{\sqrt{\dfrac{288}{23}}}))$$

5. Let consider that $\psi = \frac{1}{2}$

   With this will have $b = \frac{88}{23}$ and we have

$$\begin{aligned}
R^* =& \frac{1}{2}\phi(\frac{-\dfrac{88}{23} - \dfrac{56}{23}}{\sqrt{\dfrac{288}{23}}}) + \frac{1}{2}(1 - \phi(\frac{-\dfrac{88}{23} + (\dfrac{232}{23})}{\sqrt{\dfrac{288}{23}}}))\\
=& \frac{1}{2}\phi(\frac{-6\sqrt{46}}{23}) + \frac{1}{2}(1 - \phi(\frac{6\sqrt{46}}{23}))\\
=& \frac{1}{2}(1 - \phi(\frac{6\sqrt{46}}{23})) + \frac{1}{2}(1 - \phi(\frac{6\sqrt{46}}{23}))\\
=& (1 - \phi(\frac{6\sqrt{46}}{23}))\\
=& (1 - \phi(1.769))\\
=& (1 - 0.9615784)\\
=& 0.03842162
\end{aligned}$$