TD : Introduction au Machine Learning

1 Bayes Classifier

Consider the following mixture of Gaussian model of $(x, y) \in \mathbb{R} \times \{-1, 1\}$

$$\mathbb{P}(y = -1) = \pi_0, \quad \mathbb{P}(y = 1) = 1 - \pi_0$$
$$\mathbb{P}(x|y = -1) = \mathcal{N}(\mu_1, \sigma_1^2), \quad \mathbb{P}(x|y = 1) = \mathcal{N}(\mu_2, \sigma_2^2)$$

Question 1 : Derive the explicit formula of the conditional probability $\mathbb{P}(y = 1|x)$ and $\mathbb{P}(y = -1|x)$ in term of $\mu_1, \sigma_1, \mu_2, \sigma_2, \pi_0$. Simplify the formula in the case $\sigma_1 = \sigma_2$.

We have :

$$\begin{aligned}
\mathbb{P}(y = 1|x) &= \frac{\mathbb{P}(x|y = 1)\mathbb{P}(y = 1)}{\mathbb{P}(x)} \quad \text{(Bayes formula)} \\
&= \frac{\mathbb{P}(x|y = 1)\mathbb{P}(y = 1)}{\mathbb{P}(x \cap y = 1) + \mathbb{P}(x \cap y = -1)} \\
&= \frac{\mathbb{P}(x|y = 1)\mathbb{P}(y = 1)}{\mathbb{P}(x|y = 1)\mathbb{P}(y = 1) + \mathbb{P}(x|y = -1)\mathbb{P}(y = -1)} \\
&= \frac{\mathcal{N}(\mu_2, \sigma_2^2)(1 - \pi_0)}{\mathcal{N}(\mu_2, \sigma_2^2)(1 - \pi_0) + \mathcal{N}(\mu_1, \sigma_1^2)\pi_0} \\
&= \frac{(1 - \pi_0)\frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{(x-\mu_2)^2}{2\sigma_2^2}\right)}{(1 - \pi_0)\frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{(x-\mu_2)^2}{2\sigma_2^2}\right) + \pi_0\frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right)}
\end{aligned}$$

Similarly, we have :

$$\begin{aligned}
\mathbb{P}(y = -1|x) &= \frac{\mathbb{P}(x|y = -1)\mathbb{P}(y = -1)}{\mathbb{P}(x)} \quad \text{(Bayes formula)} \\
&= \frac{\mathbb{P}(x|y = -1)\mathbb{P}(y = -1)}{\mathbb{P}(x \cap y = 1) + \mathbb{P}(x \cap y = -1)} \\
&= \frac{\mathbb{P}(x|y = -1)\mathbb{P}(y = -1)}{\mathbb{P}(x|y = 1)\mathbb{P}(y = 1) + \mathbb{P}(x|y = -1)\mathbb{P}(y = -1)} \\
&= 1 - \mathbb{P}(y = 1|x)
\end{aligned}$$

In the case $\sigma_1 = \sigma_2$, we have :

$$\mathbb{P}(y=1|x) = \frac{(1-\pi_0)\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu_2)^2}{2\sigma^2}\right)}{(1-\pi_0)\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu_2)^2}{2\sigma^2}\right)+\pi_0\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu_1)^2}{2\sigma^2}\right)}$$
$$= \frac{(1-\pi_0)\exp\left(-\frac{(x-\mu_2)^2}{2\sigma^2}\right)}{(1-\pi_0)\exp\left(-\frac{(x-\mu_2)^2}{2\sigma^2}\right)+\pi_0\exp\left(-\frac{(x-\mu_1)^2}{2\sigma^2}\right)}$$
$$= \frac{1-\pi_0}{1-\pi_0+\pi_0\exp\left(-\frac{(x-\mu_1)^2}{2\sigma^2}+\frac{(x-\mu_2)^2}{2\sigma^2}\right)}$$

And :

$$\mathbb{P}(y=-1|x) = 1-\mathbb{P}(y=1|x)$$
$$= \frac{\pi_0\exp\left(-\frac{(x-\mu_1)^2}{2\sigma^2}+\frac{(x-\mu_2)^2}{2\sigma^2}\right)}{1-\pi_0+\pi_0\exp\left(-\frac{(x-\mu_1)^2}{2\sigma^2}+\frac{(x-\mu_2)^2}{2\sigma^2}\right)}$$

Question 2 : Give the range of $x \in \mathbb{R}$ such that $h_{\text{Bayes}}(x) = 1$. What happens to this range when $\mu_1 = \mu_2$ and $\sigma_1 = \sigma_2$?

We have :

$$h_{\text{Bayes}}(x) = 1 \Leftrightarrow \mathbb{P}(y=1|x) > \mathbb{P}(y=-1|x)$$
$$\Leftrightarrow (1-\pi_0)\frac{1}{\sqrt{2\pi\sigma_2^2}} > \pi_0\frac{1}{\sqrt{2\pi\sigma_1^2}}\exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2}+\frac{(x-\mu_2)^2}{2\sigma_2^2}\right)$$
$$\Leftrightarrow \frac{1-\pi_0}{\pi_0}\frac{\sigma_1}{\sigma_2} > \exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2}+\frac{(x-\mu_2)^2}{2\sigma_2^2}\right)$$
$$\Leftrightarrow \ln\left(\frac{1-\pi_0}{\pi_0}\frac{\sigma_1}{\sigma_2}\right) > -\frac{(x-\mu_1)^2}{2\sigma_1^2}+\frac{(x-\mu_2)^2}{2\sigma_2^2}$$
$$\Leftrightarrow \ln\left(\frac{1-\pi_0}{\pi_0}\frac{\sigma_1}{\sigma_2}\right) > \frac{x^2-2\mu_1x+\mu_1^2}{2\sigma_1^2}-\frac{x^2-2\mu_2x+\mu_2^2}{2\sigma_2^2}$$
$$\Leftrightarrow 0 > x^2\left(\frac{1}{2\sigma_1^2}-\frac{1}{2\sigma_2^2}\right)+x\left(\frac{\mu_2}{\sigma_2^2}-\frac{\mu_1}{\sigma_1^2}\right)+\frac{\mu_1^2}{2\sigma_1^2}-\frac{\mu_2^2}{2\sigma_2^2}-\ln\left(\frac{1-\pi_0}{\pi_0}\frac{\sigma_1}{\sigma_2}\right)$$

This is a quadratic inequality in x. We can solve it by finding the roots of the corresponding quadratic equation.
Then, depending on the sign of $\frac{1}{2\sigma_1^2}-\frac{1}{2\sigma_2^2}$, we can determine the range of x such that $h_{\text{Bayes}}(x) = 1$.

In the case $\mu_1 = \mu_2$ and $\sigma_1 = \sigma_2$, we have :

$$\mathbb{P}(y=1|x) = \frac{1-\pi_0}{1-\pi_0+\pi_0\exp\left(-\frac{(x-\mu_1)^2}{2\sigma^2}+\frac{(x-\mu_2)^2}{2\sigma^2}\right)} = 1-\pi_0$$
$$\mathbb{P}(y=-1|x) = \pi_0$$

Thus, $h_{\text{Bayes}}(x) = 1 \Leftrightarrow 1 - \pi_0 > \pi_0 \Leftrightarrow \pi_0 < \frac{1}{2}$.

2 Support Vector Machine

Recall the following 2 problem formulation of SVM :

$$\text{(P)} \qquad \max_{w \in \mathbb{R}^d, b \in \mathbb{R}} \frac{1}{||w||} \quad \text{s.t.} \quad \min_{i \leq m} y_i(w^T x_i + b) = 1$$

$$\text{(P')} \qquad \max_{w \in \mathbb{R}^d, b \in \mathbb{R}} \frac{1}{||w||} \quad \text{s.t.} \quad y_i(w^T x_i + b) \geq 1, \quad \forall i \leq m$$

Question 1 : Prove that if $(\bar{w}, \bar{b})$ is an optimal solution of (P), then it is an optimal solution of (P').

Let

$$E_P = \Big\{(w, b) \in \mathbb{R}^d \times \mathbb{R} \,|\, \min_{i \leq m} y_i(w^T x_i + b) = 1\Big\}$$
$$E_{P'} = \Big\{(w, b) \in \mathbb{R}^d \times \mathbb{R} \,|\, y_i(w^T x_i + b) \geq 1, \quad \forall i \leq m\Big\}$$

We have $E_P \subset E_{P'}$. Thus, if $(\bar{w}, \bar{b})$ is an optimal solution of (P), then it is an optimal solution of (P').

Consider the following linearly separable data on $\mathbb{R}^2$ with $m = 3$:

$$x_1 = (\sqrt{3}/2, 1/2), \quad x_2 = (-\sqrt{3}/2, 1/2), \quad x_3 = (0, -1)$$

The corresponding labels are $y_1 = 1, y_2 = 1, y_3 = -1$.
Question 2 : Assume $\bar{w} = (0, 4/3)$, $\bar{b} = 1/3$, compute the margin of the SMV classifier $\text{sign}(\bar{w}^T x + \bar{b})$.

We have :

$$\bar{w}^T x_1 + \bar{b} = \frac{2}{3} + \frac{1}{3} = 1$$
$$\bar{w}^T x_2 + \bar{b} = \frac{2}{3} + \frac{1}{3} = 1$$
$$\bar{w}^T x_3 + \bar{b} = \frac{-4}{3} + \frac{1}{3} = -1$$