Machine Learning Exercises: Set 1

Roger Garriga Calleja

January 30, 2017

Problem 5: Consider a binary classification problem in which the observation X is real valued, $\mathbb{P}\{Y=0\}=\mathbb{P}\{Y=1\}=\frac{1}{2}$, and the class-oriented cumulative distribution functions are

$$\mathbb{P}\{X \leqslant x|Y=0\} = \begin{cases} 0 & \text{if } x \leqslant 0 \\ \frac{x}{2} & \text{if } 0 < x \leqslant 2 \\ 1 & \text{if } x > 2 \end{cases} \quad \text{and } \mathbb{P}\{X \leqslant x|Y=1\} = \begin{cases} 0 & \text{if } x \leqslant 1 \\ \frac{(x-1)}{3} & \text{if } 1 < x \leqslant 4 \\ 1 & \text{if } x > 4 \end{cases}.$$

Determine $\eta(x)=\mathbb{P}\{Y=1|X=x\}$. Compute the Bayes classifier and the Bayes risk R^*. Compute the asymptotic risk R_{1-NN} of the nearest neighbor classifier.

Since Y is either 1 or 0, applying the law of total probabilities $\mathbb{P}(X)=\mathbb{P}(Y=1|X=x)\mathbb{P}(Y=1)+\mathbb{P}(Y=0|X=x)\mathbb{P}(Y=0)=\frac{5}{12}$. Then, using Bayes theorem

$$\eta(x)=\mathbb{P}\{Y=1|X=x\}=\frac{\mathbb{P}(X=x|Y=1)\mathbb{P}(Y=1)}{P(X)}= \tag{1}$$

$$=\frac{\mathbb{P}(X=x|Y=1)\mathbb{P}(Y=1)}{\mathbb{P}(Y=1|X=x)\mathbb{P}(Y=1)+\mathbb{P}(Y=0|X=x)\mathbb{P}(Y=0)}. \tag{2}$$

From the cdf given we can get the pdf just by taking derivatives, so

$$\mathbb{P}\{X = x|Y=0\} = \begin{cases} 0 & \text{if } x \leqslant 0 \\ \frac{1}{2} & \text{if } 0 < x \leqslant 2 \\ 1 & \text{if } x > 2 \end{cases} \quad \text{and } \mathbb{P}\{X = x|Y=1\} = \begin{cases} 0 & \text{if } x \leqslant 1 \\ \frac{1}{3} & \text{if } 1 < x \leqslant 4 \\ 1 & \text{if } x > 4 \end{cases}.$$

So we can consider only $x \in [0,4]$. Now, substituting on the equation

$$\eta(x) = \begin{cases} 0 & \text{if } x \leqslant 1 \\ \frac{2}{5} & \text{if } 1 < x < 2 \\ 1 & \text{if } 2 <\leqslant 4 \end{cases}. \tag{3}$$

$$1-\eta(x) = \begin{cases} 1 & \text{if } x \leqslant 1 \\ \frac{3}{5} & \text{if } 1 < x \leqslant 2 \\ 0 & \text{if } 2 < x \leqslant 4 \end{cases}. \tag{4}$$

The Bayes classifier takes the optimal decision as

$$p^*(X) = \begin{cases} 1 & \text{if } \eta(X) \geqslant \frac{1}{2} \\ 0 & \text{otherwise} \end{cases} \Leftrightarrow \begin{cases} 1 & \text{if } 2 < x \leqslant 4 \\ 0 & \text{otherwise} \end{cases}. \tag{5}$$

Now, we compute the Bayes risk as $R^*(x) = \mathbb{E}[\min\{\eta(x), 1-\eta(x)\}]$. The minimum is

$$\min\{\eta(x), 1-\eta(x)\} = \begin{cases} 0 & \text{if } x \leqslant 0 \\ \frac{2}{5} & \text{if } 1 < x \leqslant 2 \\ 0 & \text{if } 2 < x \leqslant 4 \end{cases}. \tag{6}$$

Finally, integrating $\min\{\eta(x), 1-\eta(x)\}\mathbb{P}(X)$ over the space $[0, 4]$ to compute the expected value we get $R^* = \frac{1}{6}$.

For the 1-NN, the asymptotic risk is computed as $R_{1-NN} = 2\mathbb{E}[\eta(x)(1-\eta(x))]$. The product is

$$\eta(x)(1-\eta(x)) = \begin{cases} 0 & \text{if } x \leqslant 1 \\ \frac{6}{25} & \text{if } 1 < x \leqslant 2 \\ 0 & \text{if } 2 < x \leqslant 4 \end{cases}. \tag{7}$$

Then, integrating $\eta(x)(1-\eta(x))\mathbb{P}(X)$ over the space $[0, 4]$ to compute the expected value we get $R_{1-NN} = \frac{1}{5}$.

Problem 6: Consider a binary classification problem in which both class-conditional densities are multivariate normal of the form

$$f_i(x) = \frac{1}{\sqrt{(2\pi)^d \det \Sigma_i}} e^{-\frac{1}{2}(x-m_i)^T \Sigma_i^{-1}(x-m_i)}, \; i = 0, 1, \tag{8}$$

where $m_i = \mathbb{E}[X|Y = i]$ and Σ_i is the covariance matrix for class i. Let $q_0 = \mathbb{P}\{Y = 0\}$ and $q_1 = \mathbb{P}\{Y = 1\}$ be the a priori probabilities.
Determine the Bayes classifier. Characterize the cases when the Bayes decision is linear (i.e, it is obtained by thresholding a linear function of x).

To determine the Bayes classifier first we need the posterior probabilities,

$$\eta(x) = \mathbb{P}(Y = 1|X = x) = \frac{f_1(x)q_1}{f_1(x)q_1 + f_0(x)q_0}, \tag{9}$$

$$1 - \eta(x) = \mathbb{P}(Y = 0|X = x) = \frac{f_0(x)q_0}{f_1(x)q_1 + f_0(x)q_0}. \tag{10}$$

Then, $p^*(x) = \begin{cases} 1 & \text{if } \eta(x) \geqslant 1-\eta(x) \\ 0 & \text{otherwise} \end{cases}. = \begin{cases} 1 & \text{if } f_1(x)q_1 \geqslant f_0(x)q_0 \\ 0 & \text{otherwise} \end{cases}$. So we need to compare $f_1(x)q_1$ with $f_0(x)q_0$ and find the values of x for which $\eta(x) \geqslant 1-\eta(x)$.

$$f_1(x)q_1 \geqslant f_0(x)q_0 \Leftrightarrow \frac{q_1}{\sqrt{\det \Sigma_1}} e^{-\frac{1}{2}(x-m_1)^T\Sigma_1^{-1}(x-m_1)} \geqslant \frac{q_0}{\sqrt{\det \Sigma_0}} e^{-\frac{1}{2}(x-m_0)^T\Sigma_0^{-1}(x-m_0)} \Leftrightarrow \tag{11}$$

$$\Leftrightarrow \log(q_1) - \frac{1}{2}\log(\det \Sigma_1) - \frac{1}{2}(x-m_1)^T\Sigma_1^{-1}(x-m_1) \geqslant \log(q_0) - \frac{1}{2}\log(\det \Sigma_0) - \frac{1}{2}(x-m_0)^T\Sigma_0^{-1}(x-m_0). \tag{12}$$

In general, the above inequation is quadratic, so the bounds would not be linear. To have a linear bound we need the second order term to be equal in both sides of the inequation, so we can get rid of it. Developing the quadratic product we get

$$(x-m_1)^T\Sigma_1^{-1}(x-m_1) \geqslant (x-m_0)^T\Sigma_0^{-1}(x-m_0) \Leftrightarrow \tag{13}$$

$$x^T\Sigma_1^{-1}x - x^T\Sigma_1^{-1}m_1 - m_1^T\Sigma_1^{-1}x + m_1^T\Sigma_1^{-1}m_1 \geqslant x^T\Sigma_0^{-1}x - x^T\Sigma_0^{-1}m_0 - m_0^T\Sigma_0^{-1}x + m_0^T\Sigma_0^{-1}m_0. \tag{14}$$

The second order term will be equal in both sides iff $\Sigma_1 = \Sigma_0$, so in this particular case the decision boundary will be linear.

Problem 7: Let the joint distribution of (X, Y) be such that X is uniform on the interval $[0, 1]$, and for all $x \in [0, 1]$, $\eta(x) = x$. Determine the prior probabilities $\mathbb{P}\{Y = 0\}$, $\mathbb{P}\{Y = 1\}$ and the class-conditional densities $f(x|Y = 0)$ and $f(x|Y = 1)$.
Calculate R^*, R_{1-NN} and R_{3-NN} (i.e., the Bayes risk and the asymptotic risk of the 1-, and 3-nearest neighbor rules).

The posterior probabilities are $\eta(x) = x$ and $1-\eta(x) = 1-x$ and the distribution of X is $p(x) = \mathbb{1}_{x\in[0,1]}$. From them we can get the prior probabilities by applying the law of total probabilities

$$\mathbb{P}(Y = 1) = \int_0^1 \mathbb{P}(Y = 1|x)p(x)\mathrm{d}x = \int_0^1 x\mathrm{d}x = \frac{1}{2} \tag{15}$$

$$\mathbb{P}(Y = 0) = 1 - \mathbb{P}(Y = 1) = \frac{1}{2}. \tag{16}$$

The class conditional densities can be calculated applying Bayes theorem

$$f(x|Y = 1) = \frac{\eta(x)\mathbb{P}(x)}{\mathbb{P}(Y = 1)} = 2x \tag{17}$$

$$f(x|Y = 0) = \frac{\eta(x)\mathbb{P}(x)}{\mathbb{P}(Y = 0)} = 2(1 - x) \tag{18}$$

Now, let's compute the risks. To compute the risk first we have to look for the minimum between $\eta(x)$ and $1 - \eta(x)$.

$$\min(\eta(x), 1 - \eta(x) = \min(x, 1 - x) = \begin{cases} x & \text{if } 0 \leqslant x \leqslant \frac{1}{2} \\ 1 - x & \text{if } \frac{1}{2} < x \leqslant 1 \end{cases}. \tag{19}$$

With that we can compute the risks:

$$R^* = \mathbb{E}[\min(\eta(x), 1 - \eta(x))] = \int_0^{\frac{1}{2}} x\mathrm{d}x + \int_{\frac{1}{2}}^1 1 - x\mathrm{d}x = \frac{1}{4}, \tag{20}$$

$$R_{1-NN} = 2\mathbb{E}[\eta(x)(1 - \eta(x))] = 2\int_0^1 x(1 - x)\mathrm{d}x = \frac{1}{3}, \tag{21}$$

$$R_{3-NN} = \mathbb{E}[\eta(x)(1 - \eta(x))] + 4\mathbb{E}[\eta(x)^2(1 - \eta(x))^2] = \frac{3}{10}. \tag{22}$$

Problem 8 Write a program that generates training data of n i.i.d. pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$ of random variables distributed such that X takes values in $\Re^d$ and $Y \in \{0, 1\}$. The join distribution is such that X is uniformly distributed in $[0, 1]^d$ and $\mathbb{P}\{Y = 1|X = x\} = x^{(1)}$ (where $x^{(1)}$ is the first component of $x = (x^{(1)}, \ldots, x^{(d)})$).
Classify X using the $1, 3, 5, 7, 9$-nearest neighbor rules. Re-draw (X, Y) many times so that you can estimate the risk of these rules. Try this for various values of n and d and plot the estimated risk. Explain what you observe.