Solution to Assignment 2

November 17, 2015

1. Given a training dataset $\mathcal{X} = \{x^{(t)}, \boldsymbol{r}^{(t)}\}_{t=1}^{N}$ where $x^{(t)} \in \mathbb{R}$ are scalars and the number of classes is $K = 2$. Suppose instances are normally distributed within each class. Write in closes-form the decision boundary $z \in \mathbb{R}$, where $P[C_1|z] = P[C_2|z]$.

   Answer:

   We want to find x such that $P[C_1|x] = P[C_2|x]$, or equivalently from Bayes' rule, $\log P[x|C_1] + \log P[C_1] = \log P[x|C_2] + \log P[C_2]$. Since instances are normally distributed within each class, we have $-\frac{1}{2}\log 2\pi - \log\sigma_1 - \frac{(x-\mu_1)^2}{2\sigma_1^2} + \log P[C_1] = -\frac{1}{2}\log 2\pi - \log\sigma_2 - \frac{(x-\mu_2)^2}{2\sigma_2^2} + \log P[C_2]$, which leads to $ax^2 + bx + c = 0$, where $a = \frac{1}{2\sigma_1^2} - \frac{1}{2\sigma_2^2}$, $b = \frac{\mu_2}{\sigma_2^2} - \frac{\mu_1}{\sigma_1^2}$, and $c = \left(\frac{\mu_1^2}{2\sigma_1^2} - \frac{\mu_2^2}{2\sigma_2^2}\right) + \log\frac{\sigma_1}{\sigma`_2} + \log\frac{P[C_2]}{P[C_1]}$. The decision boundaries are the roots of this equation, i.e., $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$.

2. Consider the univariate parametric classification. Show that the priors $P[C_i]$, $i = 1, 2, \cdots, K$, for different classes can be estimated jointly by assuming that $P[C_i]$ follows a Multinomial distribution parametrized by $\theta = (\rho_1, \cdots, \rho_K)$ with constrains $\sum_{i=1}^{K}\rho_i = 1$ and by maximizing the likelihood $P[\mathcal{X}|\theta]$.

   Answer:
   Assuming $P[C_i]$ follows a multinomial distribution parameterized by $\theta = \boldsymbol{\rho} = (\rho_1, \cdots, \rho_K)$ with constraints $\sum_{i=1}^{K}\rho_i = 1$, we wish to maximize the likelihood $P[\mathcal{X}|\theta] = \prod_{t=1}^{N}\prod_{i=1}^{K}\rho_i^{r_i^{(t)}}$.
   Define the objective as follows:

$$
\begin{aligned}
f(\boldsymbol{\rho}) &= \log(P[\mathcal{X}|\theta]) = \log\left(\prod_{t=1}^{N}\prod_{i=1}^{K}\rho_i^{r_i^{(t)}}\right) = \sum_{t=1}^{N}\sum_{i=1}^{K}\log\rho_i^{r_i^{(t)}} \\
&\quad s.t.\ g(\boldsymbol{\rho}) = \sum_{i=1}^{K}\rho_i - 1 = 0.
\end{aligned}
$$

   To solve this constraint optimization problem, we add an Lagrange multiplier α to the objective and rewrite it as

$$
L(\boldsymbol{\rho}, \alpha) = f(\boldsymbol{\rho}) - \alpha g(\boldsymbol{\rho}) = \sum_{t=1}^{N}\sum_{i=1}^{K}\log\rho_i^{r_i^{(t)}} - \alpha\left(\sum_{i=1}^{K}\rho_i - 1\right).
$$

   Taking the partial derivatives of L with respect to each ρ_i and α, we have

$$
\begin{cases}
\frac{\sum_{t=1}^{N} r_1^{(t)}}{\rho_1} & = \alpha, \\
\frac{\sum_{t=1}^{N} r_2^{(t)}}{\rho_2} & = \alpha, \\
\vdots & = \alpha, \\
\frac{\sum_{t=1}^{N} r_K^{(t)}}{\rho_K} & = \alpha, \\
\sum_{i=1}^{K} \rho_i = 1 & = 1,
\end{cases}
$$

implying that

$$\rho_i = \frac{\sum_{t=1}^{N} r_i^{(t)}}{\alpha}$$

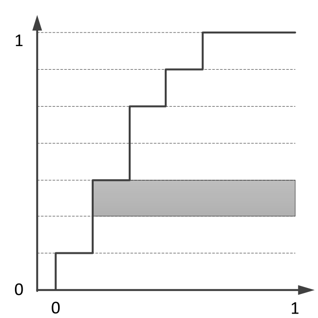
for all i. Since

$$\sum_{i=1}^{K} \rho_i = 1 = \sum_{i=1}^{K} \frac{\sum_{t=1}^{N} r_i^{(t)}}{\alpha} = \frac{\sum_{i=1}^{K} \sum_{t=1}^{N} r_i^{(t)}}{\alpha} = \frac{N}{\alpha},$$

we have $\alpha = N$.

The maximum likelihood estimator ρ_i therefore becomes $\frac{\sum_{t=1}^{N} r_i^{(t)}}{N}$.

3. Show that the *Area Under the ROC Curve* (AUC) is equal to the probability that a classifier ranks a randomly chosen positive instance higher than a randomly chosen negative one.
Answer:



Consider the horizontal partition of AUC based on each positive instance. The height of each partition (shown as the shaded area in the above figure) is the probability that a positive instance is chosen. On the other hand, the width of the partition is the conditional probability that given a positive instance is chosen, a randomly chosen negative instance is ranked after that positive instance (note that each negative instance ranked before that positive instance contributes to the portion of the horizontal bar before the shaded area). Therefore, summing up all the partition we have the joint probability that a randomly chosen positive instance is ranked before a randomly chosen negative one.