Answer Key for Practice Quiz - 1

CS 591U/791V - Pattern Recogniton
Instructor: Dr. Arun Ross
Posted on: Feb 22, 2009

1. a. Reject Option: While the goal of a classifier is to assign the input pattern (e.g., feature vector) to one of several pre-defined classes, in some instances it may be better for the classifier *not* to make a classification decision. For example, in a two-class problem, if the largest of the posterior probabilities, say $P(C_1|\boldsymbol{x})$, is relatively low, then the classifier may not assign $\boldsymbol{x}$ to either C_1 or C_2. This is called the *reject option*. Typically, when the reject option is invoked by a classifier, then human intervention may be necessary to render the final decision. Alternately, when a classifier invokes the reject option, it may be necessary to use a second (perhaps more expensive) classifier to assign the input pattern to a pre-defined class.

   b. Generative and Discriminative Models: A generative model is used to characterize the underlying class-conditional density (or the likelihood), i.e., $p(\boldsymbol{x}|C_k)$, of patterns pertaining to a specific class. Modeling the class-conditional density allows the practitioner to generate novel patterns based on the underlying model or to design a Bayes classifier by converting the class- conditional densities of multiple classes to posterior probabilities via the Bayes rule.

   A discriminative model, on the other hand, directly characterizes the posterior probability, i.e., $P(C_k|\boldsymbol{x})$, which can be used to assign the input pattern, $\boldsymbol{x}$, to that class which has the highest posterior probability.

2. The decision boundary can be computed as:

$$
\begin{aligned}
&P(C_1|x) = P(C_2|x) \\
&\Rightarrow p(x|C_1|) = p(x|C_2), \qquad \text{(since } P(C_1) = P(C_2)\text{)} \\
&\Rightarrow \frac{1}{\sqrt{2\pi 1}} exp\left(-\frac{1}{2}x^2\right) = \frac{1}{\sqrt{2\pi 4}} exp\left(-\frac{1}{2}\frac{(x-\frac{1}{2})^2}{4}\right)
\end{aligned}
$$

   Applying $\ln$ on both sides and multiplying by -1, we have:

$$
\begin{aligned}
&\frac{x^2}{2} = \ln 2 + \frac{(x-\frac{1}{2})^2}{8} \\
&\Rightarrow \frac{x^2}{2} = \ln 2 + \frac{1}{8}\left(x^2 - x + \frac{1}{4}\right) \\
&\Rightarrow 4x^2 = 8\ln 2 + (x^2 - x + \frac{1}{4}) \\
&\Rightarrow 3x^2 + x - 5.79 = 0 \\
&\Rightarrow x = -1.56 \quad \text{or} \quad x = 1.23
\end{aligned}
$$

   Thus the decision rule will be,

$$
\begin{cases} x \in C_1, & \text{if } -1.56 < x < 1.23, \\ x \in C_2, & \text{otherwise.} \end{cases}
$$

3. Let $D = \{x_1, x_2, \dots x_n\}$. Thus,

$$p(D|\theta) = p(x_1|\theta).p(x_2|\theta)\dots p(x_n|\theta) \quad \text{(since samples are iid)}$$
$$= \prod_{k=1}^{n} p(x_k|\theta)$$
$$= \prod_{k=1}^{n} \theta^{x_k}(1-\theta)^{1-x_k}$$

Applying $\ln$ on both sides:

$$\Rightarrow L = \ln P(D|\theta) = \sum_{k=1}^{n} [x_k \ln\theta + (1-x_k)\ln(1-\theta)]$$
$$\Rightarrow \frac{\partial L}{\partial \theta} = \sum_{k=1}^{n} \left[\frac{x_k}{\theta} - \frac{(1-x_k)}{(1-\theta)}\right]$$
$$= \sum_{k=1}^{n} \left[\frac{x_k(1-\theta) - (1-x_k)\theta}{\theta(1-\theta)}\right]$$
$$= \sum_{k=1}^{n} \left[\frac{x_k - \theta}{\theta(1-\theta)}\right]$$
$$\Rightarrow \frac{\partial L}{\partial \theta} = \frac{1}{\theta(1-\theta)} \sum_{k=1}^{n} [x_k - \theta]$$

To find $\hat{\theta}_{mle}$:

$$\frac{\partial L}{\partial \theta} = 0$$
$$\Rightarrow \frac{1}{\hat{\theta}_{mle}(1-\hat{\theta}_{mle})} \sum_{k=1}^{n} \left[x_k - \hat{\theta}_{mle}\right] = 0$$
$$\Rightarrow \sum_{k=1}^{n} \left[x_k - \hat{\theta}_{mle}\right] = 0$$
$$\Rightarrow \sum_{k=1}^{n} x_k - \sum_{k=1}^{n} \hat{\theta}_{mle} = 0$$
$$\Rightarrow n.\hat{\theta}_{mle} = \sum_{k=1}^{n} x_k$$
$$\Rightarrow \hat{\theta}_{mle} = \frac{1}{n}\sum_{k=1}^{n} x_k.$$
