Answer Key for Practice Exam 1

CS 591Q/791V - Pattern Recognition

Posted on: March 21, 2009

---

1. (a) Bootstrapping: A technique to estimate the variance in the error rate of a pattern classifier. In this method, the test dataset consisting of, say, N data points, $D = \{x_1, x_2, \dots x_n\}$, is sampled multiple times. Each sample set, D_S, is obtained by drawing N points at random from D, *with replacement*, so that some points in D may be replicated in D_S, whereas other points in D may be absent from D_S. This process is repeated L times to generate L sample sets each of size N. The performance of the classifier is then evaluated on each of these L sample sets thereby allowing one to estimate the variance in error between the different bootstrap data sets. [See pg. 23].

   (b) Parzen Window: The kernel function used to estimate the value of the density function at a particular point in the feature space by using the evidence of all the training pointpoints is called a Parzen window. If ϕ is a kernel function such that $\phi(\boldsymbol{u}) \geq 0$ and $\int \phi(\boldsymbol{u}) d\boldsymbol{u} = 1$, then the density at $\boldsymbol{x}$ can be estimated as,

$$\hat{p}(\boldsymbol{x}) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{h^D} \phi\left(\frac{\boldsymbol{x} - \boldsymbol{x_i}}{h}\right),$$

   where, $\{\boldsymbol{x_1}, \boldsymbol{x_2}, \dots \boldsymbol{x_N}\}$ are the training points, D is the dimensionality of each $\boldsymbol{x_i}$ and h is the width of the Parzen window. [See pg. 123].

   (c) Fisher's linear discriminant: Fisher's linear discriminant is a classification method that projects high-dimensional data ($\boldsymbol{x}$) onto a line and performs classification in this one-dimensional space (y). The projection maximizes the distance between the means of the classes while minimizing the variance within each class. The projection vector, $\boldsymbol{w}$, is obtained by maximizing the Fisher's criterion. [See pp. 186-189].

2. The estimated density at x can be computed as,

$$\hat{p}(x) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{h} \phi\left(\frac{x - x_i}{h}\right), \tag{1}$$

   where, the x_is are the points sampled from the unknown density $p(x)$, $N = 13$, $h = 1$, and ϕ is the kernel function.

   The uniform kernel function takes the following form [pg. 123]:

$$\phi(u) = \begin{cases} 1, & |u| \leq 1/2, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

   Now, $\phi\left(\frac{x - x_i}{h}\right) = \phi(x - x_i)$, $i = 1, 2, \dots 13$, since $h = 1$. Thus,

$$\phi(x - x_i) = 1, \quad \text{if } |x - x_i| \leq 1/2 \text{ based on equation (2).} \tag{3}$$

   Therefore, the sum

$$\sum_{i=1}^{13} \phi(x - x_i),$$

denotes the number of x_is that are at a distance $\leq 1/2$ from x.

So,

$$\hat{p}(0) = \frac{1}{13} \times 0 = 0,$$
$$\hat{p}(1) = \frac{1}{13} \times 2 = 0.15,$$
$$\hat{p}(3) = \frac{1}{13} \times 3 = 0.23,$$
$$\hat{p}(5) = \frac{1}{13} \times 1 = 0.07,$$
$$\hat{p}(7) = \frac{1}{13} \times 3 = 0.23,$$
$$\hat{p}(9) = \frac{1}{13} \times 0 = 0.$$

3. The required likelihood function can be written as,

$$\begin{aligned}&p\left[(x=0.6, C_1), (x=0.1, C_3), (x=0.9, C_3), (x=1.1, C_2)\right] \\ &= p(x=0.6)P(C_1|x=0.6).p(x=0.1)P(C_3|x=0.1).p(x=0.9)P(C_3|x=0.9).p(x=1.1)P(C_2|x=1.1)\end{aligned} \tag{4}$$

Note that by Bayes formula, $p(x)P(C_i|x) = p(x|C_i)P(C_i)$.

Then, equation (4) may be written as,

$$\begin{aligned}&p(x=0.6|C_1)P(C_1) \times p(x=0.1|C_3)P(C_3) \times p(x=0.9|C_3)P(C_3) \times p(x=1.1|C_2)P(C_2) \\ &= (0.33 \times 0.5) \times (0.26 \times 0.25) \times (0.39 \times 0.25) \times (0.33 \times 0.25) \\ &\text{(By evaluating the class-conditional densities at particular values of } x\text{)} \\ &= 8.62 \times 10^{-5}. \square\end{aligned}$$

4. (a) Based on the data:

$\hat{\mu_1} = 0.096$, $\hat{\sigma_1^2} = 0.657$, $\hat{\mu_2} = 0.812$, $\hat{\sigma_2^2} = 1.964$. Thus,

$$p(x|C_1) \sim N(0.096, 0.657),$$
$$p(x|C_2) \sim N(0.812, 1.964).$$

(b) The Bayes decision rule is:

Assign x to C_1 if $P(C_1|x) > P(C_2|x)$; else assign x to C_2.

If $P(C_1|x) > P(C_2|x)$, then

$$\begin{aligned}&\frac{p(x|C_1).p(C_1)}{p(x)} > \frac{p(x|C_2).p(C_2)}{p(x)} \qquad \text{(By Bayes formula)} \\ &\Rightarrow p(x|C_1) > p(x|C_2) \qquad \text{(Since } p(C_1) = p(C_2)\text{)} \\ &\Rightarrow \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left[-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right] > \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left[-\frac{(x-\mu_2)^2}{2\sigma_2^2}\right] \\ &\text{(Applying ln on both sides and multiplying both sides by } (-2)\text{)} \\ &\Rightarrow \ln(\sigma_1^2) + \frac{(x-\mu_1)^2}{\sigma_1^2} < \ln(\sigma_2^2) + \frac{(x-\mu_2)^2}{\sigma_2^2} \\ &\Rightarrow \ln(0.657) + \frac{(x-0.096)^2}{0.657} < \ln(1.964) + \frac{(x-0.812)^2}{1.964}\end{aligned}$$

$$
\begin{aligned}
&\Rightarrow -0.42 + 1.5221(x^2 - 0.192x + 0.0092) < 0.675 + 0.5092(x^2 - 1.624x + 0.6593)\\
&\Rightarrow 1.0129x^2 + 0.5348x - 1.4167 < 0\\
&\Rightarrow x < 0.95 \quad \text{and} \quad x > -1.47\\
&\Rightarrow -1.47 < x < 0.95
\end{aligned}
$$

Thus, the required classification rule is:
Assign x to C_1 if $-1.47 < x < 0.95$; else assign x to C_2. □

(c) The Bayes error rate refers to the probability of making a mistake given the Bayes decision boundary. According to the Bayes decision rule (above), x is classified as C_1 if it is in the region R_1 = (-1.47, 0.95) and C_2 if it is in the disjoint regions $R_2 = (-\infty, -1.47]$ and $[0.95, \infty)$. Thus, patterns from C_2 that fall in R_1 and patterns from C_1 that fall in R_2 will be misclassified. Therefore [see pg. 39, eqn (1.78)],

$$
\begin{aligned}
p(error) &= p(x \in R_1, C_2) + p(x \in R_2, C1)\\
&= \int_{R_1} p(x, C_2)dx + \int_{R_2} p(x, C_1)dx\\
&= \int_{R_1} p(C_2|x)p(x)dx + \int_{R_2} p(C_1|x)p(x)dx\\
&= \int_{R_1} p(x|C_2)p(C_2)dx + \int_{R_2} p(x|C_1)p(C_1)dx\\
&= 0.5\left[\int_{-1.47}^{0.95} \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left[-\frac{(x-\mu_2)^2}{2\sigma_2^2}\right] dx\right.\\
&\quad + \int_{-\infty}^{-1.47} \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left[-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right] dx\\
&\quad \left. + \int_{0.95}^{\infty} \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left[-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right] dx\right] .\square
\end{aligned}
$$

[Note: For this problem, you do not have to proceed further than this expression]
