MATH-547: Assignment # 1

Due on Monday, September 21, 2015

Saket Choudhary
2170058637

Contents

Problem 1

$(X, Y) \in R^d \times \{\pm 1\}$

$$P(Y = 1) = \pi_+$$
$$P(Y = -1) = \pi_-$$
$$P(dx|Y = 1) = p_+(x)$$
$$P(dx|Y = -1) = p_-(x)$$

Problem 1: (a)

Expression for bayes classifier:

$$P(Y = 1|x) = \frac{P(x|Y = 1)P(Y = 1)}{P(x|Y = 1)P(Y = 1) + P(x|Y = -1)P(Y = -1)}$$
$$P(Y = 1|x) = \frac{\pi_+ p_+(x)}{\pi_+ p_+(x) + \pi_- p_-(x)}$$

$$P(Y = -1|x) = \frac{P(x|Y = -1)P(Y = 1)}{P(x|Y = -1)P(Y = -1) + P(x|Y = -1)P(Y = -1)}$$
$$P(Y = -1|x) = \frac{\pi_- p_-(x)}{\pi_+ p_+(x) + \pi_- p_-(x)}$$

Define $\eta(x) = E[Y|x] = 1 \times P(Y = 1|x) + -1 \times P(Y = 1|x)$
Now,
Intuitively to minimise the error, we choose class '1' for x when $P(Y = 1|x) > P(Y = -1|X)$ and hence, the bayes classifier $g_*(x)$ is given by:

$$g_*(x) = \begin{cases} 1 & \pi_+ p_+(x) \geq \pi_- p_-(x) \\ -1 & \text{otherwise} \end{cases}$$

Where, $\pi_+ p_+(x) \geq \pi_- p_-(x)$ is given by:

$$\log(\pi_+ p_+(x)) \geq \log(\pi_- p_-(x))$$
$$\log(\pi_+) + \log(p_+(x)) \geq \log(\pi_-) + log(p_-(x))$$
$$\log(\pi_+) + \sqrt{\frac{1}{2\pi}}\sigma^{-1}(x - \vec{a_+})\sigma^{-1}(x - \vec{a_+})^T \geq \log(\pi_-) + \sqrt{\frac{1}{2\pi}}\sum^{-1}(x - \vec{a_-})\sum^{-1}(x - \vec{a_-})^T$$

Problem 1: (b)

Bayes risk: $R(f) = E[l(yf(x))]$ where $l(t) = I\{t \leq 0\}$
For 0-1 loss, the conditional risk is:
$R(i|x) = \sum_{j=0}^{k-1} L(j,i)p(j|x) \sum_{j \neq i} p(j|x) = 1 - p(i|x)$
Thus, $R(i|x)$ is the probability of x not belonging to class i
Bayes classifier found in previous part $g^*(x)$ is essentially: $g^*(x) = arg\ min_i R(i|x) = argmax_i p(i|x)$ which gave us the $\pi_+ p_+(x) \geq \pi_i p_-(x)$
Conditional risk of Bayes classifier(binary): $R(g^*(x)|x) = min(1 - p(0|x), 1 - p(1|x) = min(p(0|x), p(1|x))$
and hence the bayes risk is given by: $R^* = \int min(p(0|x), p(1|x))dF(x)$
and hence in this case, $R^* = \int min(\pi_+ p_+(x), \pi_i p_-(x))dx$

Problem 1: (c)

$$\pi_+ = \frac{\sum_{i=1}^n I\{Y_i = 1\}}{n}$$
$$\pi_- = \frac{\sum_{i=1}^n I\{Y_i = -1\}}{n}$$

We need to know $p(dx|Y = 1)$ and $p(dx|Y = -1)$ so we need estimators for a_+, a_-, σ and $\sum$
We can choose MLE estimators for multivariate guassian(derivation skipped):

$$(\hat{a_+}) = \frac{\sum I\{Y_i = 1\}}{n}$$
$$(\hat{a_-}) = \frac{\sum I\{Y_i = -1\}}{n}$$
$$\hat{\sigma} = \frac{\sum_{i;Y_i=1}(x_i - \hat{a_+})(x_i - \hat{a_+})^T}{\sum I\{Y_i = 1\}}$$
$$\hat{\sum} = \frac{\sum_{i;Y_i=-1}(x_i - \hat{a_-})(x_i - \hat{a_-})^T}{\sum I\{Y_i = 1\}}$$

And a suitable estimator of bayes classifier is :

$$\hat{g_*(x)} = \begin{cases} 1 & \hat{\pi_+}\hat{p_+(x)} \geq \hat{\pi_-}\hat{p_-(x)} \\ -1 & \text{otherwise} \end{cases}$$

where d is given by:

$$\log(\hat{\pi_+}) + \sqrt{\frac{1}{2\pi}}\hat{\sigma}^{-1}(x - \hat{\vec{a_+}})\hat{\sigma}^{-1}(x - \hat{\vec{a_+}})^T \geq \log(\hat{pi_-}) + \sqrt{\frac{1}{2\pi}}\hat{\sum}^{-1}(x - \hat{a_-})\hat{\sum}^{-1}(x - \hat{a_-})^T$$

Problem 2

Problem 2: (a)

$F(x)$ is 3 times differentiable. Consider taylor expansion of $F(x+h)$ and $F(x-h)$

$$F(x+h) = F(x) + F'(x)h + F''(x)\frac{h^2}{2} + F'''(x)\frac{h^3}{6}$$

$$F(x-h) = F(x) - F'(x)h + F''(x)\frac{h^2}{2} - F'''(x)\frac{h^3}{6}$$

Thus,

$$F(x+h) - F(x-h) = 2F'(x)h + 2F'''(x)\frac{h^3}{6}$$

$$\frac{F(x+h) - F(x-h)}{2h} = F'(x) + F'''(\epsilon)\frac{h^2}{12} \text{ for some } \epsilon \text{ in [x-h,x+h]}$$

$$|F'(x) - \frac{F(x+h) - F(x-h)}{2h}| \leq |F'''(\epsilon)\frac{h^2}{12}|$$

Problem 2 continued on next page. . .

Problem 2: (b)

Nadaraya-Watson Estimator $\eta_\eta(x) = E[Y|X = x] = \frac{\int yf(x,y)dy}{\int f(x,y)dy}$
Now,
$f(x, y) = \frac{1}{nh_xh_y}\sum_{i=1}^{n} K(\frac{x-x_i}{h_x}) \times K(\frac{y-y_i}{h_y})$

$$\int yf(x, y)dy = \frac{1}{n}\int y\sum_{i=1}^{n}\frac{1}{h_xh_y}K(\frac{x-x_i}{h_x}) \times K(\frac{y-y_i}{h_y})$$

Now, $\int y\frac{1}{h_y}K(\frac{y-y_i}{h_y})dy = y$
Hence,

$$\int yf(x, y)dy = \frac{1}{nh_x}\sum_{i=1}^{n} K(\frac{x-x_i}{)}y_i \tag{1}$$

Consider $\int f(x, y)dy$:

$$\int f(x, y)dy = \frac{1}{nh_x}\sum_{i=1}^{n} K(\frac{x-x_i}{h_x}) \times \int K(\frac{y-y_i}{h_y})dy \tag{2}$$

$$= \frac{1}{nh_x}\sum_{i=1}^{n} K(\frac{x-x_i}{h_x}) \times 1 \text{ since } \int K_{h_y}dy = 1 \tag{3}$$

Thus, using 1, 1 we get:

$$\eta_\eta(x) = \frac{\frac{1}{nh_x}\sum_{i=1}^{n} K(\frac{x-x_i}{)}y_i}{\frac{1}{nh_x}\sum_{i=1}^{n} K(\frac{x-x_i}{h_x})}$$

$$\eta_\eta(x) = \frac{\sum_{i=1}^{n} y_iK(\frac{x-x_i}{)}}{\sum_{i=1}^{n} K(\frac{x-x_i}{h_x})}$$

Problem 3

$L(g) := P(Y \neq g(X)) = E[I\{Yf(X) < 0\}]$ Consider $f(y) = max(1 - y, 0)$
$L(g) = E[I\{Yf(X) < 0\}] = E[E[max(1 - f(x), 0)|x = x]] = \int_S max(1 - f(x), 0) \times \frac{1-\eta(x)}{2} + max(1 + f(x), 0) \times \frac{1+\eta(x),2}{d}\pi$
We need to minimise the integrand: $max(1 - f(x), 0) \times \frac{1-\eta(x)}{2} + max(1 + f(x), 0) \times \frac{1+\eta(x)}{2}$
Which is to minimise: $max(1 + tf(x))(1 + t\eta(x))$ and is given by:
$f(x) = sign(\eta(x)) = g^*(x)$ [Bayes classifier]

Problem 4

Assume the sufficient condition exists, i.e.: There exist $a_1, a_2, \dots a_k \geq 0$ and binary classifiers $g_1, g_2, \dots g_k$ such that $\forall 1 \leq i \leq n$: $Y_i \sum_{j=1}^{k} a_j g_j(X_i) \geq 2\gamma$

Y_i is given by the widghted sum of predictions g_j : $Y_i = sign(\sum_{j=1}^{k} a_j g_j(X_i))$ Taking expectations:

$$E[Y_i \sum_{j=1}^{k} a_j g_j(X_i)] \geq 2\gamma$$

$$\sum_{j=1}^{k} a_j E[Y_i g_j(X_i)] \geq 2\gamma$$

Since, $\sum_j a_j = 1$ and $a_j \geq 0$ for $j = \{1, 2, \dots, k\}$ and $\sum_{j=1}^{k} a_j E[Y_i g_j(X_i)] \geq 2\gamma$ then there exists a g_j such that:

$E[Y_i g_j(X_i)] \geq 2\gamma$

$$E[Y_i g_j(X_i)] = 1 \times P[Y_i = g_j(X_i)] + -1 \times P(Y_i \neq g_j(X_i))$$

$$= 1 - 2P(Y_i \neq g_j(X_i))$$

$$\implies P(Y_i \neq g_j(X_i)) = \frac{1 - E[Y_i g_j(X_i)]}{2}$$

$$P(Y_i \neq g_j(X_i)) \leq \frac{1 - 2\gamma}{2}$$

Now for weights, $w_1, w_2, \dots w_j$ such that $\sum_j w_j = 1$:

$$\sum_{j=1}^{n} P(Y_j \neq g(X - j)) \leq \frac{1}{2} - \gamma$$

$$\sum_{j=1}^{n} E[I(Y_j \neq g(X - j))] \leq \frac{1}{2} - \gamma$$

$$\sum_{j=1}^{n} I(Y_j \neq g(X - j)) \leq \frac{1}{2} - \gamma$$