ST790 HW4

Steven Xu

October 10, 2019

1.

$$\begin{aligned}
\mathbb{E}(y_i - \hat{f}(x_i))^2 &= \mathbb{E}\left[y_i - \hat{f}(x_i) + f(x_i) - f(x_i) + \mathbb{E}\hat{f}(x_i) - \mathbb{E}\hat{f}(x_i)\right]^2 \\
&= \mathbb{E}\left[y_i - f(x_i)\right]^2 + \mathbb{E}\left[\mathbb{E}\hat{f}(x_i) - \hat{f}(x_i)\right]^2 + \mathbb{E}\left[f(x_i) - \mathbb{E}\hat{f}(x_i)\right]^2 + \\
&2\mathbb{E}\left\{\left[y_i - f(x_i)\right]\left[\mathbb{E}\hat{f}(x_i) - \hat{f}(x_i)\right]\right\} + 2\mathbb{E}\left\{\left[\mathbb{E}\hat{f}(x_i) - \hat{f}(x_i)\right]\left[f(x_i) - \mathbb{E}\hat{f}(x_i)\right]\right\} \\
&+ 2\mathbb{E}\left\{\left[y_i - f(x_i)\right]\left[f(x_i) - \mathbb{E}\hat{f}(x_i)\right]\right\} \\
\mathbf{E}_{Y^0}(y_i - \hat{f}(x_i))^2 &= \mathbb{E}\left[y_i - \hat{f}(x_i) + f(x_i) - f(x_i) + \mathbb{E}\hat{f}(x_i) - \mathbb{E}\hat{f}(x_i)\right]^2 \\
&= \mathbb{E}_{Y^0}\left[Y_i^0 - f(x_i)\right]^2 + \mathbb{E}_{Y^0}\left[\mathbb{E}\hat{f}(x_i) - \hat{f}(x_i)\right]^2 + \mathbb{E}_{Y^0}\left[f(x_i) - \mathbb{E}_{Y^0}\hat{f}(x_i)\right]^2 + \\
&2\mathbb{E}_{Y^0}\left\{\left[Y_i^0 - f(x_i)\right]\left[\mathbb{E}\hat{f}(x_i) - \hat{f}(x_i)\right]\right\} + 2\mathbb{E}_{Y^0}\left\{\left[\mathbb{E}\hat{f}(x_i) - \hat{f}(x_i)\right]\left[f(x_i) - \mathbb{E}\hat{f}(x_i)\right]\right\} \\
&+ 2\mathbb{E}_{Y^0}\left\{\left[Y_i^0 - f(x_i)\right]\left[f(x_i) - \mathbb{E}\hat{f}(x_i)\right]\right\}
\end{aligned}$$

Under the assumption $y_i, Y_i^0 = f(x_i) + \epsilon$ we have

$$\mathbb{E}\left[\mathbb{E}\hat{f}(x_i) - \hat{f}(x_i)\right]^2 + \mathbb{E}\left[f(x_i) - \mathbb{E}\hat{f}(x_i)\right]^2 + 2\mathbb{E}\left\{\left[\mathbb{E}\hat{f}(x_i) - \hat{f}(x_i)\right]\left[f(x_i) - \mathbb{E}\hat{f}(x_i)\right]\right\} =$$
$$\mathbb{E}\mathbb{E}_{Y^0}\left[\mathbb{E}\hat{f}(x_i) - \hat{f}(x_i)\right]^2 + \mathbb{E}\mathbb{E}_{Y^0}\left[f(x_i) - \mathbb{E}\hat{f}(x_i)\right]^2 + 2\mathbb{E}\mathbb{E}_{Y^0}\left\{\left[\mathbb{E}\hat{f}(x_i) - \hat{f}(x_i)\right]\left[f(x_i) - \mathbb{E}\hat{f}(x_i)\right]\right\}$$

Since $\mathbb{E}_y(\cdot) = \mathbb{E}_y\mathbb{E}_{Y^0}(\cdot)$, therefore

$$\begin{aligned}
\mathbb{E}\mathbb{E}_{Y^0}(y_i - \hat{f}(x_i))^2 - \mathbb{E}(y_i - \hat{f}(x_i))^2 = \mathbb{E}\mathbb{E}_{Y^0}\left[Y_i^0 - f(x_i)\right]^2 &+ 2\mathbb{E}\mathbb{E}_{Y^0}\left\{\left[Y_i^0 - f(x_i)\right]\left[\mathbb{E}\hat{f}(x_i) - \hat{f}(x_i)\right]\right\} + \\
&2\mathbb{E}\mathbb{E}_{Y^0}\left\{\left[Y_i^0 - f(x_i)\right]\left[f(x_i) - \mathbb{E}\hat{f}(x_i)\right]\right\} \\
&- \mathbb{E}\left[y_i - f(x_i)\right]^2 - 2\mathbb{E}\left\{\left[y_i - f(x_i)\right]\left[\mathbb{E}\hat{f}(x_i) - \hat{f}(x_i)\right]\right\} - \\
&- 2\mathbb{E}\left\{\left[y_i - f(x_i)\right]\left[f(x_i) - \mathbb{E}\hat{f}(x_i)\right]\right\}
\end{aligned}$$

Notice that

$$\mathbb{E}\mathbb{E}_{Y^0}\left[Y_i^0 - f(x_i)\right]^2 - \mathbb{E}\left[y_i - f(x_i)\right]^2 = \mathbb{E}\mathbb{E}_{Y^0}(Y_i^0)^2 - \mathbb{E}y_i^2 - 2f(x_i)\left[\mathbb{E}\mathbb{E}_{Y^0}(Y_i^0) - \mathbb{E}y_i\right] + \left[f(x_i)^2 - f(x_i)^2\right] = 0$$

$$2\mathbb{E}\mathbb{E}_{Y^0}\left\{\left[Y_i^0 - f(x_i)\right]\left[f(x_i) - \mathbb{E}\hat{f}(x_i)\right]\right\} - 2\mathbb{E}\left\{\left[y_i - f(x_i)\right]\left[f(x_i) - \mathbb{E}\hat{f}(x_i)\right]\right\} =$$

$$2\left[f(x_i) - \mathbb{E}\hat{f}(x_i)\right]\left\{\mathbb{E}\mathbb{E}_{Y^0}\left[Y_i^0 - f(x_i)\right] - \mathbb{E}\left[y_i - f(x_i)\right]\right\} = 0$$

$$\therefore \mathbb{E}\mathbb{E}_{Y^0}(y_i - \hat{f}(x_i))^2 - \mathbb{E}(y_i - \hat{f}(x_i))^2 = 2\mathbb{E}\mathbb{E}_{Y^0}\left\{\left[Y_i^0 - f(x_i)\right]\left[\mathbb{E}\hat{f}(x_i) - \hat{f}(x_i)\right]\right\} - 2\mathbb{E}\left\{\left[y_i - f(x_i)\right]\left[\mathbb{E}\hat{f}(x_i) - \hat{f}(x_i)\right]\right\}$$

Assume that Y^0 and y are independent samples, then

$$2\mathbb{E}\mathbb{E}_{Y^0}\left\{\left[Y_i^0 - f(x_i)\right]\left[\mathbb{E}\hat{f}(x_i) - \hat{f}(x_i)\right]\right\} = 2\mathbb{E}\left\{\left[\mathbb{E}\hat{f}(x_i) - \hat{f}(x_i)\right]\mathbb{E}_{Y_i^0}\left[Y_i^0 - f(x_i)\right]\right\} = 0$$

$$\therefore \mathbb{E}\mathbb{E}_{Y^0}\left[Y_i^0 - f(x_i)\right]^2 - \mathbb{E}\left[y_i - f(x_i)\right]^2 = -2\mathbb{E}\left\{\left[y_i - f(x_i)\right]\left[\mathbb{E}\hat{f}(x_i) - \hat{f}(x_i)\right]\right\}$$

$$= 2\mathbb{E}\left\{\left[y_i - f(x_i)\right]\left[\hat{f}(x_i) - \mathbb{E}\hat{f}(x_i)\right]\right\} = 2\mathrm{Cov}(y_i, \hat{y}_i)$$

$$\therefore \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\mathbb{E}_{Y^0}\left[Y_i^0 - f(x_i)\right]^2 - \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[y_i - f(x_i)\right]^2 = \frac{2}{N}\sum_{i=1}^{N}\mathrm{Cov}(y_i, \hat{y}_i)$$

2.

According the Lecture 9 Example 1, the Bayes rule under unequal cost and equal prior probabilities is

$$f^*(\mathbf{x}) = \begin{cases} 1 & \text{if } \frac{g_1(\mathbf{x})}{g_0(\mathbf{x})} > \frac{C(0,1)}{C(1,0)} \\ 0 & \text{if } \frac{g_1(\mathbf{x})}{g_0(\mathbf{x})} < \frac{C(0,1)}{C(1,0)} \end{cases}$$

$$\therefore \frac{g_1(x)}{g_0(x)} > \frac{2}{3}$$

$$\therefore \frac{\phi(x, \mu = 0, \sigma = 1)}{0.65\phi(x, \mu = 1, \sigma = 1) + 0.35\phi(x, \mu = -1, \sigma = 2)} > \frac{2}{3}$$

$$\exp\left(-\frac{x^2}{2}\right) > \frac{2}{3}\left[0.65\exp\left\{-\frac{(x-1)^2}{2}\right\} + 0.35 \times \frac{1}{4}\exp\left\{-\frac{(x+1)^2}{8}\right\}\right]$$

The decision boundary is then

$$\left\{x : \exp\left(-\frac{x^2}{2}\right) - \frac{2}{3}\left[0.65\exp\left\{-\frac{(x-1)^2}{2}\right\} + 0.35 \times \frac{1}{4}\exp\left\{-\frac{(x+1)^2}{8}\right\}\right] = 0\right\}$$

The boundary points can be found numerically using root search technique. The result is

$$\Omega_1^* = (-2.13, 1.20),\ \Omega_0^* = (-\infty, -2.13) \cup (1.20, \infty)$$

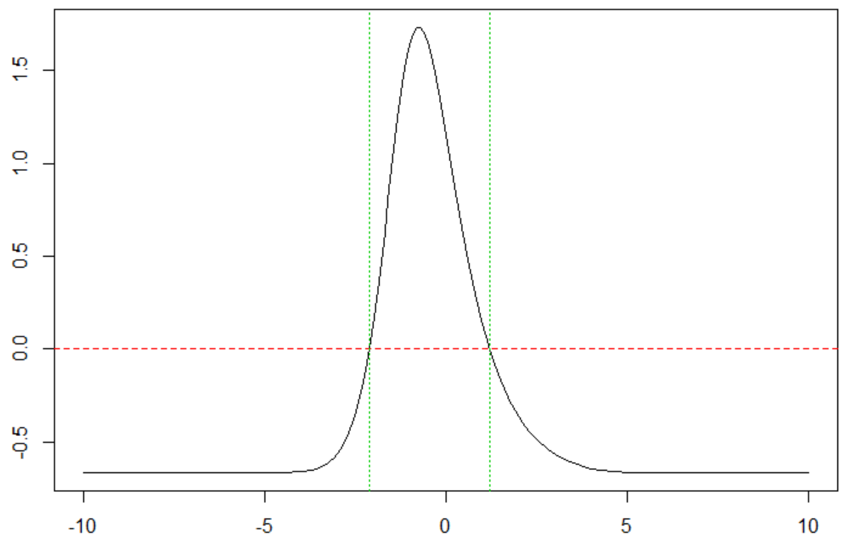

Figure 1: Decision boundary points

3.

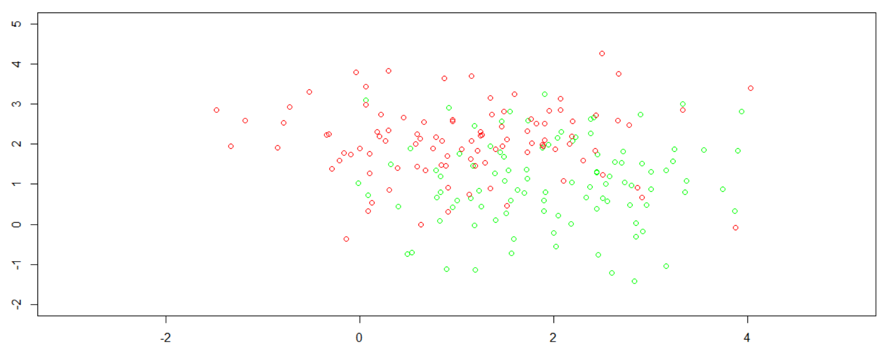

Figure 2: Training points

4.

(a)

Assuming same prior probabilities, the decision boundary under Bayes rule is

$$\{\mathbf{x} : P(Y = 1|\mathbf{X} = \mathbf{x}) = 0.5\}$$

$$\therefore \ \frac{0.5g_1(\mathbf{x})}{g(\mathbf{x})} = 0.5$$

$$\frac{g_1(\mathbf{x})}{g_0(\mathbf{x}) + g_1(\mathbf{x})} = 0.5$$

$$
\begin{aligned}
g_0(\mathbf{x}) &= g_1(\mathbf{x}) \\
(x_1 - \mu_{01})^2 + (x_2 - \mu_{02})^2 &= (x_1 - \mu_{11})^2 + (x_2 - \mu_{12})^2 \\
-2\mu_{01}x_1 + \mu_{01}^2 - 2\mu_{02}x_2 + \mu_{02}^2 &= -2\mu_{11}x_1 + \mu_{11}^2 - 2\mu_{12}x_2 + \mu_{12}^2 \\
2(\mu_{11} - \mu_{01})x_1 + 2(\mu_{12} - \mu_{02})x_2 + \mu_{01}^2 + \mu_{02}^2 - \mu_{11}^2 - \mu_{12}^2 &= 0
\end{aligned}
$$

Therefore the decision boundary is $\left\{(x_1, x_2) : x_2 = \frac{\mu_{11}^2+\mu_{12}^2-\mu_{01}^2-\mu_{02}^2-2(\mu_{11}-\mu_{01})x_1}{2(\mu_{12}-\mu_{02})}\right\}$

5.

The linear decision boundary is

$$\left\{\mathbf{x} : \mathbf{x}^T\hat{\boldsymbol{\beta}} - 0.5 = 0\right\}$$

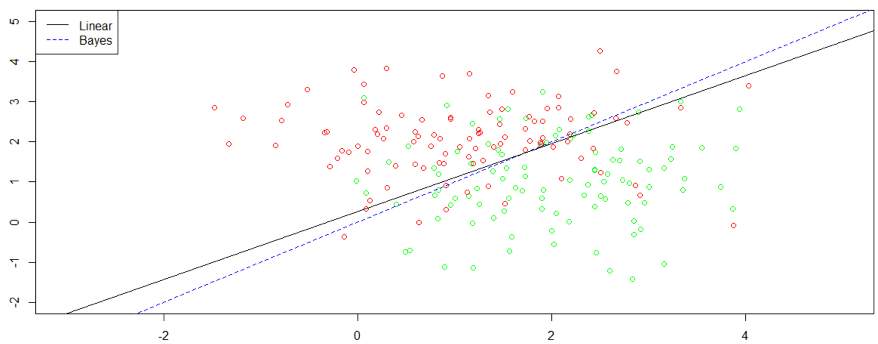

Figure 3: Linear and Bayes Decision Boundaries

Table 1: Training Error for Bayes and Linear Classifier

	False Negative	False Positive	Classification Error
Bayes	0.19	0.25	0.22
Linear	0.2	0.26	0.23

Table 2: Testing Error for Bayes and Linear Classifier

	False Negative	False Positive	Classification Error
Bayes	0.228	0.252	0.24
Linear	0.242	0.248	0.245