Math 574M Brian Mannakee
Homework 1 February 4, 2013

1. X is a continuous random variable with F(X),f(x)

   (a)

$$\begin{aligned} E(X-a)^2 &= E[X^2-2aX+a^2] \\ &= E(X^2)-2aE(X)+a^2 \end{aligned}$$

   To minimize, take the first derivative with respect to a and set equal to zero. Then take the second derivative with respect to a and check that the sign is positive.

$$\begin{aligned} \frac{d}{da}[E(X^2)-2aE(X)+a^2] &= 0 \\ 2E(X)+2a &= 0 \\ a &= E(X) \\ \frac{d}{da}[2E(X)+2a] &= 2 \end{aligned}$$

   Therefore $a = E(X)$ is a minimum

   (b) Set the derivative equal to zero

$$\begin{aligned} E(|X-a|) &= \int |X-a| f(x)dx \\ &= \int_{-\infty}^{a} -(X-a)f(x)dx + \int_{a}^{\infty} (X-a)f(x)dx \\ \frac{d}{da}E(|X-a|) &= \frac{d}{da}[\int_{-\infty}^{a} -(X-a)f(x)dx + \int_{a}^{\infty} (X-a)f(x)dx] \\ 0 &= \int_{-\infty}^{a} f(x)dx - \int_{a}^{\infty} f(x)dx \\ \int_{-\infty}^{a} f(x)dx &= \int_{a}^{\infty} f(x)dx \end{aligned}$$

   This equality only holds when a is the median of the distribution. In addition $d^2/da^2 = 2f(a) > 0$ so the median is the minimum.

2. If $X_1,\ldots,X_n$ are i.i.d Bernoulli(p) then $Y = \sum X_i \sim$ Binomial(n,p). If $p \sim beta(\alpha,\beta)$ then their joint pdf is

$$\begin{aligned} f(y,p) &= \left[\binom{n}{y} p^y (1-p)^{n-y}\right]\left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(beta)} p^{\alpha-1}(1-p)^{\beta-1}\right] \\ &= \binom{n}{y}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{y+\alpha-1}(1-p)^{n-y+\beta-1} \end{aligned}$$

   We get the marginal of y by integrating out p from the joint distribution.

$$\begin{aligned} f(y) &= \int_0^1 \binom{n}{y}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{y+\alpha-1}(1-p)^{n-y+\beta-1}dp \\ &= \binom{n}{y}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 p^{y+\alpha-1}(1-p)^{n-y+\beta-1}dp \\ &= \binom{n}{y}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(y+\alpha)\Gamma(n-y+\beta)}{\Gamma(n+\alpha+\beta)} \text{By definition of the Beta function} \end{aligned}$$

The posterior $f(p|y)$ is given by the ratio of the joint and the marginal $\frac{f(y,p)}{f(y)}$

$$f(p|y) = \frac{\Gamma(n+\alpha+\beta)}{\Gamma(y+\alpha)\Gamma(n-y+\beta)} p^{y+\alpha-1}(1-p)^{n-y+\beta-1} \sim beta(y+\alpha, n-y+\beta)$$

(a) The posterior mean is the mean of $beta(y+\alpha, n-y+\beta)$ which is

$$\hat{p} = \frac{y+\alpha}{\alpha+\beta+n} = \frac{\sum_{i=1}^n X_i + \alpha}{\alpha+\beta+n} //$$

(b) Decompose MSE $= E(p-\hat{p})^2$ into bias and variance:

$$\begin{aligned} Var_p(\hat{p}) + Bias_p(\hat{p}) &= Var_p\left(\frac{\frac{\sum_{i=1}^n X_i+\alpha}{\alpha+\beta+n} + \alpha}{\alpha+\beta+n}\right) + \left(E_p\left(\frac{\sum_{i=1}^n X_i + \alpha}{\alpha+\beta+n}\right) - p\right)^2 \\ &= \frac{1}{(\alpha+\beta+n)^2} Var_p\left(\sum_{i=1}^n X_i\right) + \left(\frac{E_p\left(\sum_{i=1}^n X_i\right)+\alpha}{\alpha+\beta+n} - p\right)^2 \\ &= \frac{np(1-p)}{\alpha+\beta+n)^2} + \left(\frac{np+\alpha}{\alpha+\beta+n} - p\right)^2 \text{ because} \sum_{i=1}^n X_i \sim \text{binomial(n,p)} \end{aligned}$$

(c) Let $\alpha = \beta = \sqrt{n/4}$ then

$$\hat{p} = \frac{\sum_{i=1}^n X_i + \sqrt{n/4}}{\sqrt{n/4} + \sqrt{n/4} + n} = \frac{\sum_{i=1}^n X_i + \sqrt{n/4}}{\sqrt{n} + n} //$$

$$\begin{aligned} R(p,\hat{p}) &= \frac{np(1-p)}{\alpha+\beta+n)^2} + \left(\frac{np+\alpha}{\alpha+\beta+n} - p\right)^2 \\ &= \frac{np-np^2}{(\alpha+\beta+n)^2} + \left(\frac{np+\alpha-p\alpha-p\beta-np}{(\alpha+\beta+n)}\right)^2 \\ &= \frac{np-np^2}{(\alpha+\beta+n)^2} + \frac{(\alpha-p(\alpha+\beta))^2}{(\alpha+\beta+n)^2} \\ &= \frac{np-np^2+\alpha^2-2p\alpha(\alpha+\beta)+p^2(\alpha+\beta)^2}{(\alpha+\beta+n)^2} \\ &= \frac{np-np^2+\frac{n}{4}-np+np^2}{(n+\sqrt{n})^2} \\ &= \frac{n}{4(n+\sqrt{n})^2} // \end{aligned}$$

3. (a) $m(x) = \int_\Theta f(x,\theta)d\theta = \int_\Theta f(x|\theta)\pi(\theta)d\theta$

(b)

$$
\begin{aligned}
r_B(\pi,\hat{\theta}(\mathbf{X})) &= \int_\Theta R(\theta,\hat{\theta})\pi(\theta)d\theta \\
&= \int_\Theta \int_\mathcal{X} L(\theta,\hat{\theta}((X))\pi(\mathbf{x}|\theta)d\mathbf{x}\pi(\theta)d\theta \text{ Definition of R()} \\
&= \int_\Theta \int_\mathcal{X} L(\theta,\hat{\theta}((X))f(\theta,\mathbf{x})d\mathbf{x}d\theta \text{ Bayes Rule} \\
&= \int_\mathcal{X} \left[\int_\Theta L(\theta,\hat{\theta}((X))\pi(\theta|\mathbf{x})d\theta\right] m(\mathbf{x})d\mathbf{x} \text{ Bayes Rule} \\
&= \int_\mathcal{X} r(\hat{\theta}|\mathbf{x})m(\mathbf{x})d\mathbf{x} \text{ Definition of } r(\hat{\theta}|\mathbf{x})//
\end{aligned}
$$

(c) The Bayes' Rule $\hat{\theta}(vecx)$ minimizes the posterior risk $r(\hat{\theta}|\mathbf{x}) = \int_\Theta L(\theta,\hat{\theta})\pi(\theta|\mathbf{x})d\theta$, so we take its derivative under squared error loss and set it equal to 0

$$
\begin{aligned}
\frac{d}{d\hat{\theta}(\mathbf{x})} r(\hat{\theta}|\mathbf{x}) &= \frac{d}{d\hat{\theta}(\mathbf{x})} \int_\Theta (\theta - \hat{\theta}(\mathbf{x})^2\pi(\theta|\mathbf{x})d\theta \\
0 &= -2\int_\Theta \theta\pi(\theta|\mathbf{x})d\theta + 2\int_\Theta \hat{\theta}(\mathbf{x})\pi(\theta|\mathbf{x})d\theta \\
\int_\Theta \theta\pi(\theta|\mathbf{x})d\theta &= \hat{\theta}(\mathbf{x})\int_\Theta \pi(\theta|\mathbf{x})d\theta = \hat{\theta}(\mathbf{x}) * 1 \\
E(\theta|\mathbf{X}=\mathbf{x}) &= \hat{\theta}(\mathbf{x})//
\end{aligned}
$$

4. See prob1.r

5. (a) Scenario 1 is a balanced classification problem, so the Bayes' boundary can be written as

$$\left\{\mathbf{x} : \frac{g_1(\mathbf{x})}{g_0(\mathbf{x})} = 1\right\}$$

The two classes are generated by the bivariate normal processes outlined in question 4 giving

$$\left\{\mathbf{x} : \frac{(2\pi)^{-1}\exp\{-\frac{1}{2}[(x_1-2)^2+(x_2-1)^2]\}}{(2\pi)^{-1}\exp\{-\frac{1}{2}[(x_1-1)^2+(x_2-2)^2]\}} = 1\right\}$$

$$\left\{\mathbf{x} : \frac{(2\pi)^{-1}\exp\{-\frac{1}{2}[x_1^2-4x_1+4+x_2^2-2x_2+1]\}}{(2\pi)^{-1}\exp\{-\frac{1}{2}[x_1^2-2X_1+1+x_2^2-4x_2+4]\}} = 1\right\}$$

$$\left\{\mathbf{x} : x_1^2-4x_1+4+x_2^2-2x_2+1 = x_1^2-2x_1+1+x_2^2-4x_2+4\right\}$$

$$\{\mathbf{x} : x_1 = x_2\}$$

So the Bayes Rule gives a line with intercept 0 and slope 1, above which points belong to class one and below which they belong to class 0. A plot with the Bayes' Rule line in blue is at the end of this document. The training error for the Bayes Rule is 29%. The testing error for the Bayes Rule is 22.8%.

(b) We can construct the Bayes Rule for Scenario 2 in the same way, although since the means are unknown it can't be simplified meaningfully

$$\left\{ \mathbf{x} : \frac{(2\pi)^{-1} \exp\{-\frac{5}{2}[(x_1 - \mu_{k1})^2 + (x_2 - \mu_{k2})^2]\}}{(2\pi)^{-1} \exp\{-\frac{5}{2}[(x_1 - \nu_{k1})^2 + (x_2 - \nu_{k2})^2]\}} = 1 \right\}$$

6. (a) For Scenario 1 the linear estimator derived from the test data is $\hat{C}(\mathbf{x}) = .5762 + .1606X_1 - .2039X_2$ and we will use the decision rule

$$\hat{f}(\mathbf{x}) = \begin{cases} 1 & \text{if } \hat{C}(\mathbf{x}) > 0.5 \\ 0 & \text{if } \hat{C}(\mathbf{x}) < 0.5 \end{cases}$$

A plot with the linear model decision boundary plotted as a blue dashed line is included at the bottom of this document. The training error for the linear model is 25%, and the testing error is 23.2%

(b) For Scenario 2 the linear estimator derived from the test data is $\hat{C}(\mathbf{x}) = .36129 + .27892X_1 - .06001X_2$ and we will use the same decision rule. The training error for the linear model is 18.5%, and the testing error is 19%

Plots Blue solid line is the Bayes boundary. Dashed blue lines are the linear model decision boundary.