Problem Set 4

Key (with contributions from E. Peairs and P. Stallworth)

March 11, 2016

4. (a) Note carefully the language: *within 10% of the range of X closest to that test observation.* That's a bit amgibuous. It's helpful that they give the example that suggests that the total range should be 10%. So it's a fair interpretation to think that at $X = 0$ we'd have from $[0, 0.1]$ to use in making our prediction.

   The fraction of observations available for prediction is proportional to the proportion of the range of X utilizd by the model. Here, it will be 10%. It's possible to interpret their language differently here, in which case you might come up with an answer somewhat lower than 10%.

   (b) Now we're using the squared percentage from last time, or $0.1^2 = 0.01 = 1\%$. This represents the proportion of the unit square covered by a square with a side of length 0.1.

   (c) Now we're using $0.1^100 = 1e - 100$ of the data.

   (d) We can see from this that each time you add a predictor, you reduce the number of observations by a factor of the fractional window size, which gets really tiny for large p.

   (e) Since we normalized all the data to the range $[0, 1]$, the volume of the data space is always 1. Thus, we always want to contain a volume of 0.1. If l is the length of the cube's side, our requirement is that $l^p = 0.1$, so $l = 0.1^{1/p}$. This yields $l = 0.1, 0.316,$ and 0.978 for $p = 1, 2,$ and 100, respectively.

6. (a) Plugging in the provided values to the logistic function, $p = \dfrac{1}{1 + e^{6-0.05\cdot 40-1\cdot 3.5}} = 0.38$

   (b) Varying X_1 and holding X_2 fixed, we need to solve

$$
\begin{aligned}
0.5 &= \frac{1}{1 + e^{6-0.05\cdot X_1-3.5}} \\
\implies 0.5 + 0.5e^{2.5-0.05\cdot X_1} &= 1 \\
\implies e^{2.5-0.05\cdot X_1} &= 1 \\
\implies 2.5 &= 0.005X_1 \\
\implies X_1 &= 50
\end{aligned}
$$

7. **Mathematical approach**: We have that $\pi_{yes} = 0.8$, and also that $f_{yes}(x) = \frac{1}{\sqrt{2\pi\cdot 36}}e^{-(x-10)^2/2\cdot 36} = \frac{1}{\sqrt{72\pi}}e^{-(x-10)^2/72}$. Plugging into Bayes' theorem, we can calculate our probability to be

$$
\frac{0.8\cdot\frac{1}{\sqrt{72\pi}}e^{-(x-10)^2/72}}{0.8\cdot\frac{1}{\sqrt{72\pi}}e^{-(x-10)^2/72} + 0.2\cdot\frac{1}{\sqrt{72\pi}}e^{-(x)^2/72}} = \frac{1}{1 + \frac{1}{4}e^{(-20x+100)/72}}
$$

Plugging in $X = 4$ gives us

$$p(4) = \frac{1}{1 + \frac{1}{4}e^{(-80+100)/72}} = 0.7518$$

Computational approach: From the problem, we (approximately) learn $X|(Y = \text{No}) \sim N(0, 34)$ and $X|(Y = \text{No}) \sim N(10, 34)$. Furthermore, we know that $\mathbb{E}[Y] = P(Y = 1) = 0.8$. I use the following `R` code to determine $P(X|Y = \text{Yes})$ and $P(X|Y = \text{No})$.

```
pxNo <- dnorm(4, mean = 0, sd = sqrt(36))
pxYes <- dnorm(4, mean = 10, sd = sqrt(36))
```

and now we do the full bayes theorem:

$$P(Y = 1|X) = \frac{P(X|Y = 1)P(Y = 1)}{P(X|Y = 1)P(Y = 1) + P(X|Y = 0)P(Y = 0)}$$

```
(0.8 * pxYes)/(0.8 * pxYes + 0.2 * pxNo)

## [1] 0.7518525
```

So $P(Y = \text{Yes}|X = 4) = 0.752$.