CFRM 521 Homework 2

Evan Hofmeister

April 14, 2018

Question 1A) - See R line 1-50

I used a local polynomial regression to predict the 5-day move in the future using the loess and predict R functions. I found the sum of squares of errors in the predictions to be: 147,640.8

Question 1B)

In part B, I found the sum of squares of errors in the predictions to be: 357,537.7 As you can see the local polynomial regression is has less than half the sum of squared error than the prediction in model B. This indicates that even though the local polynomial regression is not great at predicitng, it is much better than using something like the 5 day increment used in part B to predict, so we can see that this type of modeling in part A has some value.

Question 2A)

part A) We can tell that this function $s(x)$ is a PDF because it meets the criteria for a PDF and not a CDF. The area under its curve formes a triangle that sums to 1 and this function is always non-negative. Because this area has a finite area under the curve, and because the other criteria are met, it is a PDF. A function cannot be a PDF and CDF as it would contradict the criteria given above.

Part B) From part A, we know x is a PDF so we can find:

$$0.1 = \int_y^2 (2-x)dx$$
$$= 4 - 2 - 2y + \frac{y^2}{2}$$
$$= \frac{y^2}{2} - 2y + 2$$
$$next:$$
$$y = 2 \pm \frac{(2^2 - 4(\frac{1}{2})(1.9))}{2(\frac{1}{2})}$$

$$.1 = \frac{(2-y)^2}{2}$$

$$y = \pm .2^{\frac{1}{2}}$$

We know the solution must be between $\in (1,2)$ so $y = 2 - .2^{\frac{1}{2}}$

You could also find the area under the curve by finding the area under the triangle the function formed, that I mentioned between $x \in (1,2)$

Question 2B)

From the given equation we can find:

$$S(1.5) = 1 + .65(1.5) + 1.5^2 = .5^2 = 4.475$$

$$S'(x) = .65 + 2x + 2(x-1)_+ + 1.2(x-2)_+$$
$$S'(1.5) = .65 + 2x + 2(.5) + 1.2(0) = 4.65$$

$$S''(x) = 2 + 2(x-1)_+^0 + 1.2(x-1.2)_+^0$$
$$S''(2.2) = 2 + 2(1.0) + 1.2(1.0) = 5.2$$

Question 3A)

Class conditional probabilities for words given in the order of the problem:

Sports: 5/7 5/7 2/7 5/7 2/7 1/7 1/7 1/7 politics: 2/6 1/6 1/6 5/6 5/6 1/6 4/6 5/6

Based on the data:

$$p(politics) = \frac{6}{13} = 0.462 \quad p(sport) = \frac{7}{13} = 0.538$$

From the given testing data document x, contains goal, defence, offence, wicket and office. From this we can find:

$$p(\text{x}|sport) = \frac{5}{7} \cdot \frac{2}{7} \cdot \frac{5}{7} \cdot \frac{5}{7} \cdot \frac{2}{7} \cdot \frac{1}{7} \cdot \frac{1}{7} \cdot \frac{6}{7} = \frac{3000}{5764801} = 0.00052$$

$$p(\text{x}|politics) = \frac{2}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{1}{6} \cdot \frac{4}{6} \cdot \frac{1}{6} = \frac{5000}{1679616} = 0.002977$$

Then we can find our goal $\Rightarrow$

$$p(\text{politics }|\text{x}) = \frac{p(\text{politics}) \cdot p(\text{x}|\text{politics})}{p(\text{politics}) \cdot (\text{x}|\text{politics}) + p(\text{sport}) \cdot p(\text{x}|\text{sport})} = 0.831$$

Which is the probability that this document, x, is about politics.

Question 3B)

• To find the MLE estimators of the parameters, we first find the log likelihood.

$$L := \log\left(\prod_{i=1}^{N} p_{\mathrm{X},C}(\mathrm{X}_n,\ C_n)\right)$$

$$= \sum_{n=1}^{N} \log p_{\mathrm{X},C}(\mathrm{X}_n,\ C_n) N = \sum_{n=1}^{N} \log\left(N(p_C(C_n)\prod_{i=1}^{D} p_{X_i|C}(X_{n,i}|C_n))\right)$$

$$= \sum_{n=1}^{N} \log(p_C(C_n)) + \sum_{n=1}^{N}\sum_{i=1}^{D} \log(p_{X_i|C}(X_{n,i}|C_n))$$

$$= \sum_{n=1}^{N} \log(p_C(1))1_{\{C_n=1\}} + \sum_{n=1}^{N} \log(p_C(2))1_{\{C_n=2\}} + \sum_{n=1}^{N} \log(p_C(0))1_{\{C_n=0\}}$$

$$+ \sum_{n=1}^{N}\sum_{i=1}^{D} \log(\theta_{i,C_n}^{X_{i,n}}(1-\theta_{i,C_n})^{1-X_{i,n}})$$

$$= \sum_{n=1}^{N} \log(p_C(1))1_{\{C_n=1\}} + \sum_{n=1}^{N} \log(p_C(2))1_{\{C_n=2\}} + \sum_{n=1}^{N} \log(1-p_C(1)-p_C(2))1_{\{C_n=0\}}$$

$$+ \sum_{n=1}^{N}\sum_{i=1}^{D} \log(\theta_{i,C_n}^{X_{i,n}}(1-\theta_{i,C_n})^{1-X_{i,n}})$$

Note that $p_C(0) + p_C(1) + p_C(2) = 1$

Let us define:

$$n_2 = \sum_{n=1}^{N} 1_{\{c_n=2\}}$$

$$n_1 = \sum_{n=1}^{N} 1_{\{c_n=1\}}$$

$$n_0 = \sum_{n=1}^{N} 1_{\{c_n=0\}}$$

Next we can derive:

$$L = \log(p_C(1))\sum_{n=1}^{N} 1_{\{C_n=1\}} + \log(p_C(2))\sum_{n=1}^{N} 1_{\{C_n=2\}} + \log(1-p_C(1)-p_C(2))\sum_{n=1}^{N} 1_{\{C_n=0\}}$$

$$+ \sum_{n=1}^{N}\sum_{i=1}^{D} [X_{i,n}\log\theta_{i,C_n} + (1-X_{i,n})\log(1-\theta_{i,C_n})]$$

$$= n_1\log p_C(1) + n_2\log p_C(2) + n_0\log(1-p_C(1)-p_C(2)) + \sum_{n=1}^{N}\sum_{i=1}^{D}\sum_{c=0}^{2} 1_{\{X_{i,n}=1,C_n=c\}}\log\theta_{x',c}$$

$$+ \sum_{n=1}^{N}\sum_{i=1}^{D}\sum_{c=0}^{2} 1_{\{X_{i,n}=0,C_n=c\}}\log(1-\theta_{i,c})$$

$$= n_1\log p_C(1) + n_2\log p_C(2) + n_0\log(1-p_C(1)-p_C(2))$$

$$+ \sum_{i=1}^{D}\sum_{c=0}^{2} [n_{1,i,c}\log\theta_{i,c} + n_{2,i,c}\log\theta_{i,c} + n_{0,i,c}\log(1-\theta_{i,c})]$$

Note: we defined $n_{x,i,c} = \sum_{n=1}^{N} 1_{\{X_{i,n}=x,C_n=c\}}$ (the number of observations of class $C=c$ for which $X_i = x$.)

- The MLE estimators of $p_C(1)$ and $\theta_{i,c}$ are the maximizers of L. From the first order derivative test, we have

$$\frac{\partial L}{\partial p_C(1)} = \frac{n_1}{\hat{p}_C(1)} - \frac{n_0}{1 - \hat{p}_C(1) - \hat{p}_C(2)} = 0 \Rightarrow \hat{p}_C(1) = \frac{n_1(1 - \hat{p}_C(2))}{n_0 + n_1}$$

similarly:

$$\frac{\partial L}{\partial p_C(2)} = \frac{n_2}{\hat{p}_C(2)} - \frac{n_0}{1 - \hat{p}_C(1) - \hat{p}_C(2)} = 0 \Rightarrow \hat{p}_C(2) = \frac{n_2(1 - \hat{p}_C(1))}{n_0 + n_2}$$

We can solve this system of equations for $p_C(1)$ and $p_C(2)$:

$$\hat{p}_c(2) = \frac{n_2(n_0 + n_1) - n_2 n_1}{(n_0 + n_2)(n_0 + n_1) - n_2 n_1} \Rightarrow \hat{p}_C(2) = \frac{n_2}{n_0 + n_1 + n_2}$$

from this we can solve for: $\hat{p}_C(1) = \dfrac{n_1}{n_0 + n_1 + n_2}$

next we find theta (as defined by the professor in office hours):

$$\frac{\partial L}{\partial \theta_{i,c}} = \frac{n_{1,i,c}}{\hat{\theta}_{i,c}} - \frac{n_{0,i,c}}{1 - \hat{\theta}_{i,c}} = 0 \Rightarrow \hat{\theta}_{i,c}(1) = \frac{n_{1,i,c}}{n_{1,i,c} + n_{0,i,c}} = \frac{\text{number of observations of class C with Xi} = 1}{\text{total number if observations of class C}}$$

given new observations for the predictor variables, $X^* = (X_1^*, \ldots, X_D^*)$, we need to find class C^*.

To find $C^*|X^*$, use the Bayes' rule:

$$p_{C^*|X^*}(c|x) = \frac{p_{X^*|C} * (x|c) p_{C^*}(c)}{p_{X^*}(x)}.$$

To classify the observation, Use the following rule: We assign a class $C^* = 0$ if $p_{C^*|X} * (0|x) > p_{C^*|X} * (1|x)$ and $p_{C^*|X} * (0|x) > p_{C^*|X} * (2|x)$

We assign a class $C^* = 1$ if $p_{C^*|X} * (1|x) > p_{C^*|X} * (0|x)$ and $p_{C^*|X} * (1|x) > p_{C^*|X} * (2|x)$

We assign a class $C^* = 2$ if $p_{C^*|X} * (2|x) > p_{C^*|X} * (1|x)$ and $p_{C^*|X} * (2|x) > p_{C^*|X} * (0|x)$

For the case $p_{C^*|X} * (1|x) = p_{C^*|X} * (0|x)$, we arbitrarily choose to assign $C^* = 0$. For the case $p_{C^*|X} * (2|x) = p_{C^*|X} * (0|x)$, we arbitrarily choose to assign $C^* = 0$.

Rule 1: The classification rule (for $C^* = 0$) can be written as follows:

$$\frac{p_{X^*|C^*}(x|0) p_{C^*}(0)}{p_{X^*}(x)} > \frac{p_{X^*|C^*}(x|1) p_{C^*}(1)}{p_{X^*}(x)}$$

$$\Rightarrow$$

$$p_{X^*|C^*}(x|0) p_{C^*}(0) > p_{X^*|C^*}(x|1) p_{C^*}(1) \Rightarrow$$

$$p_C(0) \prod_{i=1}^{D} \theta_{i,0}^{X_i}(1 - \theta_{i,0})^{1-X_i} > p_C(1) \prod_{i=1}^{D} \theta_{i,1}^{X_i}(1 - \theta_{i,1})^{1-X_i} \Rightarrow$$

$$\log\left(\frac{p_C(0)}{p_C(1)}\right) + \sum X_i^* \log\left(\frac{\theta_{i,0}(1 - \theta_{i,0})}{\theta_{i,1}(1 - \theta_{i,1})}\right) > 0.$$

next we need to look at the case where:

$$\frac{p_{X^*|C^*}(x|0) p_{C^*}(0)}{p_{X^*}(x)} > \frac{p_{X^*|C^*}(x|2) p_{C^*}(2)}{p_{X^*}(x)}$$

$$\Rightarrow$$

$$p_{X^*|C^*}(x|0) p_{C^*}(0) > p_{X^*|C^*}(x|2) p_{C^*}(2) \Rightarrow$$

4

$$p_C(0) \prod_{i=2}^{D} \theta_{i,0}^{X_i} (2 - \theta_{x',0})^{2-X_i} > p_C(2) \prod_{i=2}^{D} \theta_{i,2}^{X_i} (2 - \theta_{i,2})^{2-X_i} \Rightarrow$$

$$\log\left(\frac{p_C(0)}{p_C(2)}\right) + \sum X_i^* \log\left(\frac{\theta_{i,0}(2 - \theta_{i,0})}{\theta_{i,2}(2 - \theta_{i,2})}\right) > 0$$

Thus, the classification rules can be written as if $a + \sum_{i=1}^{D} \omega_i X_i^* > 0$ AND $b + \sum_{i=1}^{D} \tau_i X_i^* > 0$, for constants a, b, ω_i and τ_i.

The boundaries $a + \sum_{i=0}^{D} \omega_i X_i^* = 0$ and $b + \sum_{i=1}^{D} \tau_i X_i^* = 0$ are hyperplane in the predictor space B^D which means, observation are classified as $c = 0$ if they are on the "positive" side of the hyperplanes.

Rule 2: The classification rule (for $C^* = 1$) can be written as follows:

$$\frac{p_{X^*|C^*}(x|1)p_{C^*}(1)}{p_{X^*}(x)} > \frac{p_{X^*|C^*}(x|0)p_{C^*}(0)}{p_{X^*}(x)}$$

$$\Rightarrow$$

$$p_{X^*|C^*}(x|1)p_{C^*}(1) > p_{X^*|C^*}(x|0)p_{C^*}(0) \Rightarrow$$

$$p_C(1) \prod_{i=0}^{D} \theta_{i,1}^{X_i} (1 - \theta_{i,1})^{0-X_i} > p_C(0) \prod_{i=0}^{D} \theta_{i,0}^{X_i} (1 - \theta_{i,0})^{0-X_i} \Rightarrow$$

$$\log\left(\frac{p_C(1)}{p_C(0)}\right) + \sum X_i^* \log\left(\frac{\theta_{i,1}(1 - \theta_{i,1})}{\theta_{i,0}(1 - \theta_{i,0})}\right) > 1$$

next we need to look at the case where:

$$\frac{p_{X^*|C^*}(x|1)p_{C^*}(1)}{p_{X^*}(x)} > \frac{p_{X^*|C^*}(x|2)p_{C^*}(2)}{p_{X^*}(x)}$$

$$\Rightarrow$$

$$p_{X^*|C^*}(x|1)p_{C^*}(1) > p_{X^*|C^*}(x|2)p_{C^*}(2) \Rightarrow$$

$$p_C(1) \prod_{i=2}^{D} \theta_{i,1}^{X_i} (1 - \theta_{i,1})^{2-X_i} > p_C(2) \prod_{i=2}^{D} \theta_{i,2}^{X_i} (1 - \theta_{i,2})^{2-X_i} \Rightarrow$$

$$\log\left(\frac{p_C(1)}{p_C(2)}\right) + \sum X_i^* \log\left(\frac{\theta_{i,1}(1 - \theta_{i,1})}{\theta_{i,2}(1 - \theta_{i,2})}\right) > 0$$

Thus, the classification rules can be written as if $q + \sum_{i=0}^{D} \zeta_i X_i^* > 0$ AND $r + \sum_{i=0}^{D} \phi_i X_i^* > 0$, for constants q, r, ζ_i and ϕ_i.

The boundaries $q + \sum_{i=1}^{D} \zeta_i X_i^* = 0$ and $r + \sum_{i=0}^{D} \phi_i X_i^* = 0$ are hyperplane in the predictor space B^D which means, observation are classified as $c = 1$ if they are on the "positive" side of the hyperplanes.

Rule 3:

The classification rule (for $C^* = 1$) can be written as follows:

$$\frac{p_{X^*|C^*}(x|2)p_{C^*}(2)}{p_{X^*}(x)} > \frac{p_{X^*|C^*}(x|0)p_{C^*}(0)}{p_{X^*}(x)}$$

$$\Rightarrow$$

$$p_{X^*|C^*}(x|2)p_{C^*}(2) > p_{X^*|C^*}(x|0)p_{C^*}(0) \Rightarrow$$

$$p_C(2)\prod_{i=0}^{D}\theta_{i,2}^{X_i}(1-\theta_{i,2})^{0-X_i} > p_C(0)\prod_{i=0}^{D}\theta_{i,0}^{X_i}(1-\theta_{i,0})^{0-X_i} \Rightarrow$$

$$\log\left(\frac{p_C(2)}{p_C(0)}\right) + \sum X_i^* \log\left(\frac{\theta_{i,2}(1-\theta_{i,2})}{\theta_{i,0}(1-\theta_{i,0})}\right) > 0$$

next we need to look at the case where:

$$\frac{p_{X^*|C^*}(x|2)p_{C^*}(2)}{p_{X^*}(x)} > \frac{p_{X^*|C^*}(x|1)p_{C^*}(1)}{p_{X^*}(x)}$$

$$\Rightarrow$$

$$p_{X^*|C^*}(x|2)p_{C^*}(2) > p_{X^*|C^*}(x|1)p_{C^*}(1) \Rightarrow$$

$$p_C(2)\prod_{i=1}^{D}\theta_{i,2}^{X_i}(1-\theta_{i,2})^{1-X_i} > p_C(1)\prod_{i=1}^{D}\theta_{i,1}^{X_i}(1-\theta_{i,1})^{1-X_i} \Rightarrow$$

$$\log\left(\frac{p_C(2)}{p_C(1)}\right) + \sum X_i^* \log\left(\frac{\theta_{i,2}(1-\theta_{i,2})}{\theta_{i,1}(1-\theta_{i,1})}\right) > 0$$

Thus, the classification rules can be written as if $t + \sum_{i=0}^{D}\alpha_i X_i^* > 0$ AND $u + \sum_{i=0}^{D}\beta_i X_i^* > 0$, for constants t, u, α_i and β_i.

The boundaries $t + \sum_{i=2}^{D}\alpha i X_i^* = 0$ and $u + \sum_{i=0}^{D}\beta_i X_i^* = 0$ are hyperplane in the predictor space B^D which means, observation are classified as $c = 2$ if they are on the "positive" side of the hyperplanes.