STAT 391

Final Exam Solutions

2:30 – 4:20 pm on June 8, 2018

©Marina Meilă
mmp@stat.washington.edu

Problem 1.

1.1.1 Comparing A with B, we see the KDE in A has visibly more variation than that of B, and B is more "averaged out". Hence B eigher has more data or has larger bandwidth. Here only the choice "same data, $h_B > h_A$" applies.

1.1.2 Looking at the behaviour of the KDE on the domains $x < -2$ and $x > 0$, and notice that C is smoother than A. If C has the same amount of data, it is impossible that these patterns are not shown in plot A. If C has less data than A, then A must has way more data to "mask" the patterns, and even so would have way bigger bandwidth. Hence naturally we guess that C has more data than A. Here only the choice "More data $h_C = h_A$" applies.

1.2 A has less bias than B

1.3 C has less variance than A

Problem 2.

$$P(X = k) = \frac{\lambda^k}{k!}e^{-\lambda},$$

$$P(X = 0 \text{ or } 1) = \frac{\lambda^0}{0!}e^{-\lambda} + \frac{\lambda^1}{1!}e^{-\lambda} = (1 + \lambda)e^{-\lambda}.$$

By definition of likelihood we have $L(\lambda|\mathcal{D}) = \prod_{i=1}^n L(\lambda|x_i)$, where $L(\lambda|x_i = k > 1) = \lambda^k e^{-\lambda}$ (omitting the constant), and $L(\lambda|x_1 = 0 \text{ or } 1) = (1 + \lambda)e^{-\lambda}$. Let $n_k = \#\{x_i : x_i = k\}$ the number of observations that has outcome k,

$$l(\lambda|\mathcal{D}) = n_{01}\left[\ln(1 + \lambda) - \lambda\right] + \sum_{k=2}^{\infty} n_k \left[k \ln \lambda - \lambda\right] = -n\lambda + n_{01} \ln(1 + \lambda) + \ln \lambda \sum_{k=2}^{\infty} k n_k$$

If we were to put in numbers,

$$l(\lambda|\mathcal{D}) = -20\lambda + 8 \ln(1 + \lambda) + 32 \ln \lambda.$$

Problem 3.

- Can be all. Anything on the bottom of the line would has closer neighour being the circle, the logistic and linear coincides in this simple case, and the quadratic has a special case of being linear in this case.
- Same as above
- Not 1NN: the decision boundary is not half-way between points. QDA will be a curved boundary. LDA is possible
- 1NN has polygon boundary, not possible to be circle. Logistic and linear both have linear boundary. Quadratic is possible

Problem 4.

1 count the number of free parameters, 0,1,2.

2 (Same as from hw7) $b_1^{ML} = b_2^{ML} = \max_i x_i$, $a_2^{ML} = \min_i x_i$.

3

$$l_0 = \log(1^n) = 0, l_1 = \log([\frac{1}{b}]^n) = -n \log b, l_2 = \log([\frac{1}{b-a}]^n) = -n \log(b-a)$$

4

$$\mathrm{BIC}_0 = 0 - 0 = 0, \mathrm{BIC}_1 = -n \log b - \log(n)/2, \mathrm{BIC}_2 = -n \log(b-a) - 2\log(n)/2$$

5 Now

$$\mathrm{BIC}_0 = 0 - 0 = 0, \mathrm{BIC}_1 = -16 \log_2 b - 2, \mathrm{BIC}_2 = -16 \log_2(b-a) - 4$$

To prefer $\mathcal{M}_0$, need $\mathrm{BIC}_0 > \mathrm{BIC}_1$, i.e., $0 > -16\log_2(b) - 2 \Rightarrow b > 2^{-1/8}$.

6 To prefer $\mathcal{M}_2$, need $\mathrm{BIC}_2 > \mathrm{BIC}_1$, i.e., $-16\log_2(a-b) - 4 > -16\log_2(b) - 2 \Rightarrow \frac{b}{a-b} > 2^{-1/8}$.

7

$$P(\mathrm{BIC}_0 > \mathrm{BIC}_1) = P(b > 2^{-1/8}) = 1 - P(x_1, \ldots, x_n < 2^{-1/8}) = 1 - \prod_{i=1}^{16} P(X_i < 2^{-1/8}) = 1 - [2^{-1/8}]^{16} = 0.75.$$

Problem 5.

1 $L(y_i|x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{1}{2\sigma^2}(y_i/\beta - x_i)^2)$.

2 $l(\beta : \sigma^2|\mathcal{D}) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^n \left(\frac{y_i^2}{\beta^2} - \frac{2x_i y_i}{\beta} + x_i^2\right)$.

3 Take derivative wrt β, $\frac{\partial}{\partial\beta} l = -\frac{2}{2\sigma^2}\sum_{i=1}^n \left(\frac{x_i y_i}{\beta^2} - \frac{y_i^2}{\beta^3}\right) = 0$ hence $\beta^{ML} = \frac{\sum_{i=1}^n y_i^2}{\sum_{i=1}^n x_i y_i}$.

4 Given X_i's, the sufficient statistics are $\sum_{i=1}^n x_i y_i, \sum_{i=1}^n y_i^2$.

5 *Too hard, no simple solution. Removed from consideration.*