STA314 Homework 6

student number: 1003942326

Yulin WANG

3/12/2019

Question 1

(a)

Since $x \leq \xi$, so $(x-\xi)_+^3 = 0$, then

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$$

Thus, $f(x) = f_1(x)$ for all $x \leq \xi$ when

$$a_1 = \beta_0;\ b_1 = \beta_1;\ c_1 = \beta_2;\ d_1 = \beta_3$$

(b)

Since $x > \xi$, so $(x-\xi)_+^3 = (x-\xi)^3 = x^3 - 3x^2\xi + 3x\xi^2 - \xi^3$, then

$$\begin{aligned} f(x) &= \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta 4(x-\xi)^3 \\ &= \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta 4(x^3 - 3x^2\xi + 3x\xi^2 - \xi^3) \\ &= (\beta_0 - \beta_4\xi^3) + (\beta_1 + 3\beta_4\xi^2)x + (\beta_2 - 3\beta_4\xi)x^2 + (\beta_3 + \beta_4)x^3 \end{aligned}$$

Thus, $f(x) = f_2(x)$ for all $x > \xi$ when

$$a_2 = \beta_0 - \beta_4\xi^3;\ b_2 = \beta_1 + 3\beta_4\xi^2;\ c_2 = \beta_2 - 3\beta_4\xi;\ d_2 = \beta_3 + \beta_4$$

(c)

$$\begin{aligned} f_1(\xi) &= \beta_0 + \beta_1\xi + \beta_2\xi^2 + \beta_3\xi^3 \\ f_2(\xi) &= (\beta_0 - \beta_4\xi^3) + (\beta_1 + 3\beta_4\xi^2)\xi + (\beta_2 - 3\beta_4\xi)\xi^2 + (\beta_3 + \beta_4)\xi^3 \\ &= \beta_0 + \beta_1\xi + \beta_2\xi^2 + (\beta_4 + 3\beta_4 - 3\beta_4 + \beta_3 + \beta_4)\xi^3 \\ &= \beta_0 + \beta_1\xi + \beta_2\xi^2 + \beta_3\xi^3 \end{aligned}$$

Thus, we've proved $f_1(\xi) = f_2(\xi)$, that is $f(x)$ is continuous at ξ.

(d)

$$\begin{aligned} f_1'(x) &= b_1 + 2c_1 x + 3d_1 x^2 \\ &= \beta_1 + 2\beta_2 x + 3\beta_3 x^2 \\ \Rightarrow f_1'(\xi) &= \beta_1 + 2\beta_2\xi + 3\beta_3\xi^2 \end{aligned}$$

$$\begin{aligned} f_2'(x) &= b_2 + 2c_2 x + 3d_2 x^2 \\ &= (\beta_1 + 3\beta_4\xi^2) + 2(\beta_2 - 3\beta_4\xi)x + 3(\beta_3 + \beta_4)x^2 \\ \Rightarrow f_2'(\xi) &= \beta_1 + 3\beta_4\xi^2 + 2\beta_2\xi - 6\beta_4\xi^2 + 3(\beta_3 + \beta_4)\xi^2 \\ &= \beta_1 + 2\beta_2\xi + (3\beta_4 - 6\beta_4 + 3\beta_3 + 3\beta_4)\xi^2 \\ &= \beta_1 + 2\beta_2\xi + 3\beta_3\xi^2 \end{aligned}$$

Thus, we've proved $f_1'(\xi) = f_2'(\xi)$, that is $f'(x)$ is continuous at ξ.

(e)

$$\begin{aligned} f_1''(x) &= 2c_1 + 6d_1x \\ &= 2\beta_2 + 6\beta_3 x \\ \Rightarrow f_1''(\xi) &= 2\beta_2 + 6\beta_3\xi \end{aligned}$$

$$\begin{aligned} f_2''(x) &= 2c_2 + 6d_2x \\ &= 2(\beta_2 - 3\beta_4\xi) + 6(\beta_3 + \beta_4)x \\ \Rightarrow f_2''(\xi) &= 2\beta_2 - 6\beta_4\xi + 6(\beta_3 + \beta_4)\xi \\ &= 2\beta_2 + (-6\beta_4 + 6\beta_3 + 6\beta_4)\xi \\ &= 2\beta_2 + 6\beta_3\xi \end{aligned}$$

Thus, we've proved $f_1''(\xi) = f_2''(\xi)$, that is $f''(x)$ is continuous at ξ.

Question 2

(a)

As $\lambda \to 0$, $\hat{g}(2)$ will have the smaller training RSS since it will be a higher order polynomial due to the order of the penalty term (it will be more flexible).

(b)

As $\lambda \to 0$, $\hat{g}(1)$ will have the smaller test RSS since $\hat{g}(2)$ is more flexible and it may cause overfitting.

(c)

For $\lambda = 0$, $\hat{g}(1) = \hat{g}(2)$, so they will have the same training and test RSS.

Question 3

(a) Give an equation for each measure

Gini index:

$$G = \hat{P}_{m1}(1 - \hat{P}_{m1}) + \hat{P}_{m2}(1 - \hat{P}_{m2}) = 2\hat{P}_{m1}(1 - \hat{P}_{m1})$$

classification error:

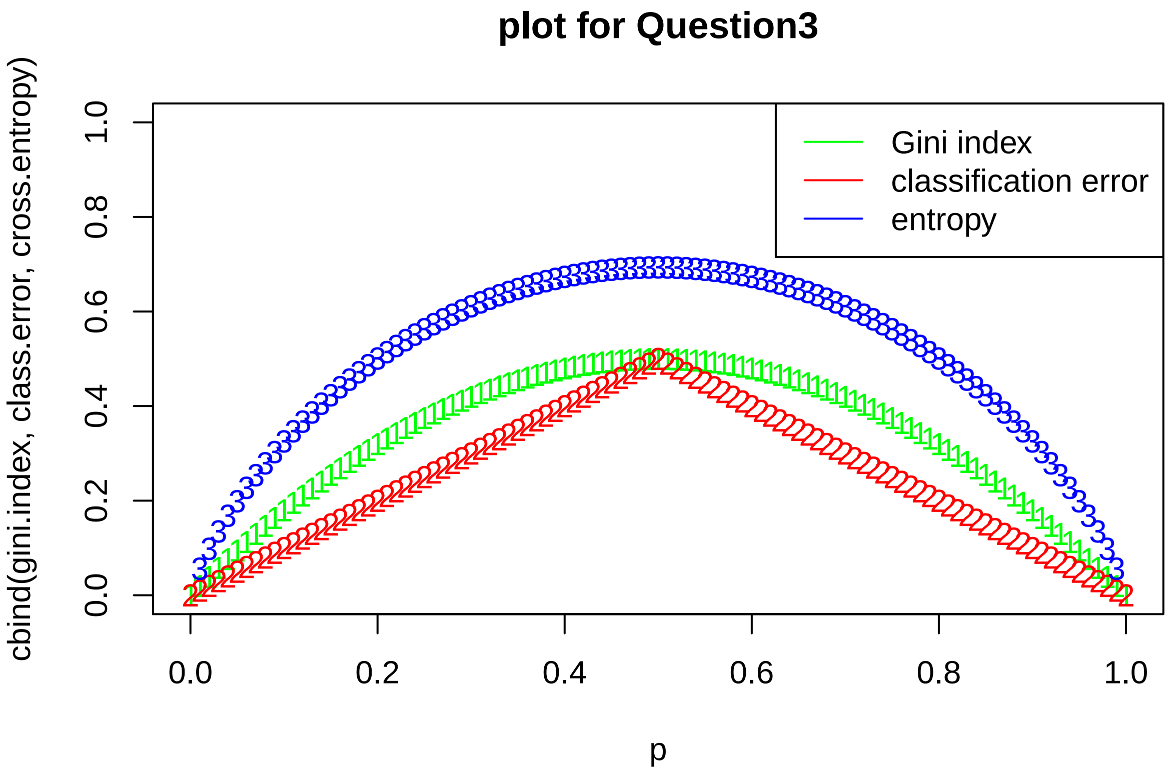
$$E = 1 - max(\hat{P}_{m1}, \hat{P}_{m2}) = 1 - max(\hat{P}_{m1}, 1 - \hat{P}_{m1})$$

entropy:

$$D = -\hat{P}_{m1}log(\hat{P}_{m1}) - \hat{P}_{m2}log(\hat{P}_{m2}) = -\hat{P}_{m1}log(\hat{P}_{m1}) - (1 - \hat{P}_{m1})log(1 - \hat{P}_{m1})$$

(b) Plot

```r
p <- seq(0, 1, 0.01)
gini.index <- 2 * p * (1 - p)
class.error <- 1 - pmax(p, 1 - p)
cross.entropy <- - p * log(p) - (1 - p) * log(1 - p)
matplot(p, cbind(gini.index, class.error, cross.entropy),
        col = c("green", "red", "blue"), ylim = c(0,1),
        main = "plot for Question3")
legend("topright", legend = c("Gini index", "classification error", "entropy"),
       col = c( "green", "red", "blue"), lty = 1)
```



Question 4

(a)

With the majority vote approach, we classify X as Red.
Because it is the most commonly occurring class among the 10 predictions (6 for Red vs 4 for Green).

(b)

With the average probability approach, we classify X as Green.
Because the average of the 10 probabilities is 0.45, which is smaller than 0.5.