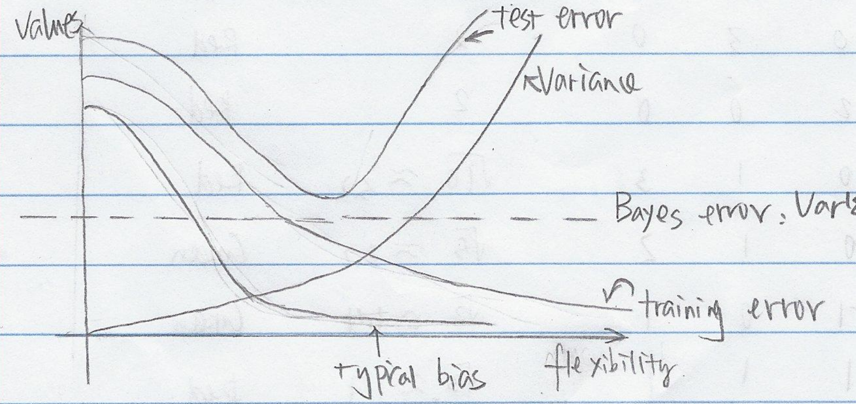
Jiayuan Shi
MA 685 HW1

Exercise 1

Chap2. 3a.



3b. All the five lines are nonnegative

①. Typical bias curve decreases monotonically because it increases in flexibility yield a closer fit

②. Variance curve increases monotonically because it increases in flexibility yield overfit

③. Training error curve decreases monotonically because it increases in flexibility yield a closer fit

④. Test error curve exhibits a characteristic U-shape, because it increases in flexibility yields a closer fit before it overfits and excessive flexible

⑤. Bayes error curve is the horizontal dashed line, indicating Var(ε), which corresponds to the lowest achievable test MSE among all possible methods. When the training error is lower than the Bayes error, overfitting has taken place.

The Bayes error rate is defined for classification problems and is determind by the ratio of data points which lie at the "wrong" side of the decision boundary ($0 \le$ value < 1)

7a.

Obs	X_1	X_2	X_3	Distance to (0,0,0)	Y
1	0	3	0	3	Red
2	2	0	0	2	Red
3	0	1	3	$\sqrt{10} \approx 3.2$	Red
4	0	1	2	$\sqrt{5} \approx 2.2$	Green
5	-1	0	1	$\sqrt{2} \approx 1.4$	Green
6	1	1	1	$\sqrt{3} \approx 1.7$	Red.

7b. Green. The observarion #5 is the closest neighbor for $k=1$

7c. Red. The observation #2, 5, 6 are the closest neighbors for $k=3$. #2 is Red, #5 is Green and #6 is Red.

7d. Small. When k is small, the decision boundary is inflexible and find patterns in the data that don't correspond to the Bayes decision boundary. As k grows, the method becomes less flexible and produces a decision boundary that is close to linear.

Exercise 2.

i. (a). The relationship between the predictors and the response is not linear

(b). There is an outlier which is far from the value predicted by the model

(c). There is an observation has high leverage, in that the predictor value for this observation is large relative to the other observations.

(d). The variance of error term is not constant. The funnel shape indicates heteroscedasticity

(e). From the plot of residuals versus predicted values, a strong curve pattern in the residuals indicates non-linearity in the data

(f). The error terms $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ in the plot are not uncorrelated. It shows an increasing pattern between $\hat{\varepsilon}_{t-1}$ and $\hat{\varepsilon}_t$

1. ~~The Ho~~

ii. • For TV: in the presence of radio and newspaper adversiting, TV adversiting does not affect sales.

• For radio: H: In the presence of TV and newspaper adversiting, radio adversiting does not affect sales.

• For newspaper: in the presence of TV and radio adversiting, newspaper adversiting does not affect sales.

Because the p-values of TV and radio < 0.0001, we can reject ~~the~~ their null hypotheses. So TV and radio adversiting can affect sales. The p-value of newspaper is 0.8599, we fail to reject its Ho., and newspaper adversiting does not affect sales.

3a. $\underset{Y}{salary} = 50 + 20 \cdot \underset{X_1}{gpa} + 0.07 \cdot \underset{X_2}{iq} + 35 \cdot \underset{X_3}{gender} + 0.01 \cdot \underset{X_1}{gpa} \cdot \underset{X_2}{iq} - 10 \cdot \underset{X_1}{gpa} \cdot \underset{X_3}{gender}$

When $X_3 = 0$ (Male)

$$Y = 50 + 20X_1 + 0.07X_2 + 0.01X_1 \cdot X_2$$

When $X_3 = 1$ (Female)

$$Y = 50 + 20X_1 + 0.07X_2 + 35 + 0.01X_1X_2 - 10X_1$$

⇒ When GPA is high enough, males earn more on average

⇒ iii is correct

3b. Gender = 1, IQ = 110, GPA = 4.0

$$salary = 50 + 20 \cdot 4 + 0.07 \cdot 110 + 35 + 0.01 \cdot 4 \cdot 110 - 10 \cdot 4 = 137.1$$

3c. False. We need to see the p-value of the regression coefficient to know if the interaction is significant or not.