Homework5 Solution

Question2: Suppose we have a dataset with 3 features: X_1 = GPA, X_2 = Age, X_3 = Type of Position (1 for Technical positions, and 0 for Non-Technical positions), and we have built a non-linear regression model as:

$$\text{Target} = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \theta_3 X_3 + \theta_4 X_1 X_2 + \theta_5 X_1 X_3$$

The prediction target is "starting salary after graduation" (in thousands of dollars). Suppose we train (fit) the model, and get θ_0 = 30, θ_1 = 20, θ_2 = 0.07, θ_3 = -30, θ_4 = 0.01, θ_5 = 10.

(a) Which answer is correct, and why?

i. For a fixed value of Age and GPA, Technical positions earn more on average than non-technical positions.

ii. For a fixed value of Age and GPA, Non-Technical positions earn more on average than Technical positions.

iii. For a fixed value of Age and GPA, Technical positions earn more on average than Non-Technical positions when the GPA is high enough.

iv. For a fixed value of Age and GPA, Non-Technical positions earn more on average than Technical positions when the GPA is high enough.

(b) Predict the salary of a Technical and a Non-Technical positions with Age of 27, GPA of 4.0.

Answer:

(a) Here is the above regression model after replacing the coefficients:

$$Salary = 30 + 20X_1 + 0.07X_2 - 30X_3 + 0.01X_1X_2 + 10X_1X_3$$

For non-technical positions, X_3=0, so the 4th term of the equation ($-30X_3$), and 6th term of the above equation ($10X_1X_3$) are zero.
So, for non-technical positions the salary will be:

$$Salary = 30 + 20X_1 + 0.07X_2 + 0.01X_1X_2$$

However, for technical positions, X_3=1, so the 4th term in the equation is -30, and 6th term is $10X_1$. So, for technical positions the salary will be:

$$Salary = 30 + 20X_1 + 0.07X_2 + 0.01X_1X_2 + 10X_1 - 30$$

Now, it all depends on X_1: If X_1 (GPA) is more than 3, then $10X_1 - 30 > 0$, thus the total salary of technical positions is higher than non-technical. But, If X_1 (GPA) is less than 3, then $10X_1 - 30 < 0$, thus the total salary of technical positions is less than non-technical.

Thus, the correct answer is (iii)

(b)

$$Salary.non.technical = 30 + 20 * 4 + 0.07 * 27 + 0.01 * 4 * 27 = 112.97K\ \$$$

$$Salary.technical = 30 + 20 * 4 + 0.07 * 27 + 0.01 * 4 * 27 + 10 * 4 - 30 = 122.97K\ \$$$