Quiz 3: Introduction to Data science: Tools and Techniques **Total Marks: 9** 2019-04-22

Name: ---------- **Registration #: ----------** **Section: ------**

Q1. If we plot the residual, which of the following options indicate that there is margin of improvement in the model?

a) clear pattern in plot b) healthy plot c) Heteroscedasticity in plot d) Randomness in residuals

Q2. Let the coefficient of determination to be computed is 0.89 in a problem involving one independent variable (x) and one dependent variable (y). The results means that

a) The relationship between two variables is negative

b) The correlation coefficient is 0.89 also

c) 89% of the total variation is explained by the independent variable

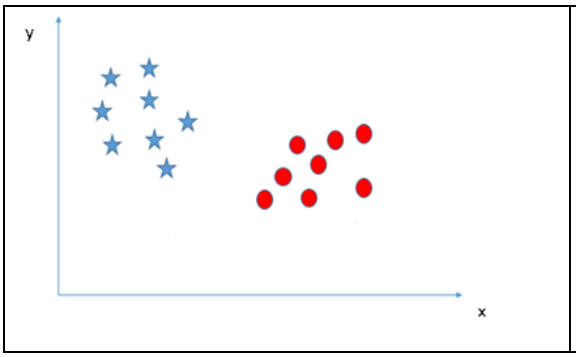
d) 89% of the total variation is explained by the dependent variable

Q3. [1 mark] EDA - Data Transformation: Suppose the relationship between X (independent variable) and Y (dependent variable) is represented by a function **Y =** $4x^2$. Your task is to transform (natural log transformation) the data in such a way that we can fit a Linear Line using Linear Regression. What will be the approximated intercept term and slope of the line.

Intercept: ---------- **Slope: ----------**

Q4. When we use SVM, can we find a local optimal solution? (Yes/No) Explain your choice of selection.

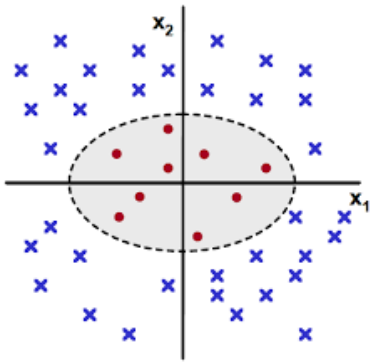
Q5. [1 marks] Discuss SVM and Logistic Regression with respect to decision boundary for the following dataset. Draw the decision boundary for both.



Quiz 3: Introduction to Data science: Tools and Techniques **Total Marks: 9** 2019-04-22

Name: ---------- **Registration #: ----------** **Section: ------**

Q6. [1 mark] Suppose you have the following dataset with 2 features. How can we use SVM to classify this data?



Solution:

Q7. [1 mark] Suppose you fit a linear regression model on the given dataset, how would you decide if there is any room for improvement in the model or not?

Q8. [2 Marks] A friend of yours is faced with a regression problem with two possible inputs, X_1 and X_2. he/she considers a linear regression model: $h(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2$

You train the model using gradient descent and the trained parameters are $\theta_0 = 3$, $\theta_1 = 2$, $\theta_2 = 4$.

The data set is given in the following table:

x_1	x_2	Y (output)	h(x)
2	2	11	
3	3	23	
1	4	21	
5	3	27	

Your task is to compute the coefficient of determination R^2 and tell how good the model fits the data. where $R^2 = 1 - (SE_{line} / SE_{\bar{y}})$