# Day 2 Quiz (Regression Analysis)
## Machine Learning and Data Analytics using Python

Electronics & ICT Academy, IIT Roorkee
09th September-13th September, 2019

Name of the participant
…….

---

**Q1.** Which of the following assumptions do we make while deriving linear regression parameters?

1. The true relationship between dependent y and predictor x is linear
2. The model errors are statistically independent
3. The errors are normally distributed with a 0 mean and constant standard deviation
4. The predictor x is non-stochastic and is measured error-free

A. 1,2 and 3
B. 1,3 and 4
C. 1 and 3
D. All of the above

**Q2.** To test linear relationship of y(dependent) and x(independent) continuous variables, which of the following plot best suited?

A. Scatter plot
B. Bar chart
C. Histograms
D. None of these

**Q3.** Generally, which of the following method(s) can be used as a classifier?

1. Linear Regression
2. Logistic Regression

A. 1 and 2
B. only 2
C. only 1
D. None of these.

**Q4.** The following visualization shows the fit of three different models (in blue line) on same training data. What can you conclude from these visualizations?

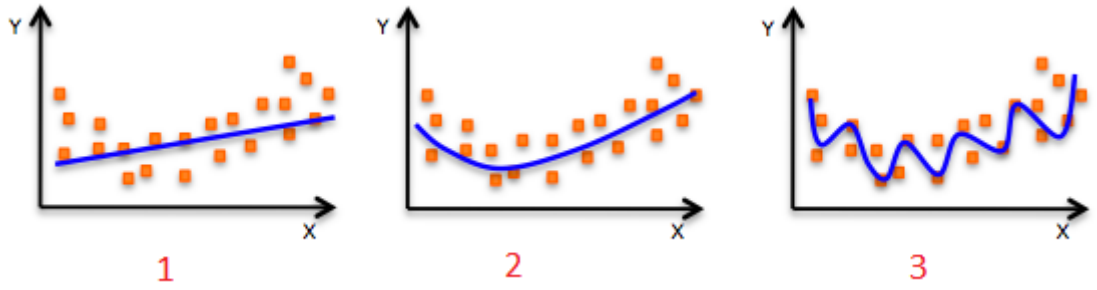


*Figure 1. fit of three different models.*

1. The training error in first model is higher when compared to second and third model.
2. The best model for this regression problem is the last (third) model, because it has minimum training error.
3. The second model is more robust than first and third because it will perform better on unseen data.
4. The third model is overfitting data as compared to first and second model.
5. All models will perform same because we have not seen the test data.

A. 1 and 3
B. 1 and 3
C. 1, 3 and 4
D. Only 5

**Q5.** If two variables are correlated, is it necessary that they have a linear relationship?

A. Yes
B. No
C. They may or may not
D. They would have non-linear relationship with 0.5 probability.

**Q6.** In a simple linear regression model (One independent variable), If we change the input variable by 1 unit. How much output variable will change?

A. A: By 1
B. No change
C. By intercept
D. By its Slope

**Q7.** Now we increase the training set size gradually. As the training set size increases, what do you expect will happen with the mean training error?

A. Increase
B. Decrease
C. Remain
D. Can't Say

**Q8**. Suppose that you have a dataset D1 and you design a linear regression model of degree 3 polynomial and you found that the training and testing error is "0" or in another terms it perfectly fits the data. What will happen when you fit degree 4 polynomial in linear regression? What will happen when you fit degree 2 polynomial in linear regression?