

## **Day 2 Quiz (Regression Analysis) - Solutions**

### **Machine Learning and Data Analytics using Python**

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

---

#### **Solution 1: D**

**Explanation:** When deriving regression parameters, we make all four assumptions mentioned above. If any of the assumptions is violated, the model would be misleading.

#### **Solution 2: A**

**Explanation:** To test the linear relationship between continuous variables, scatter plot is a good option. We can find out how one variable is changing w.r.t. another variable. A scatter plot displays the relationship between two quantitative variables.

#### **Solution 3: B**

**Explanation:** Generally, Logistic Regression is used as a classifier. However, by using threshold values along with regression, Linear Regression can also be used as a classifier.

#### **Solution 4: C**

**Explanation:** The trend of the data looks like a quadratic trend over independent variable X. A higher degree (Right graph) polynomial might have a very high accuracy on the train population but is expected to fail badly on test dataset. But if you see in left graph we will have training error maximum because it under-fits the training data.

#### **Solution 5: B**

**Explanation:** It is not necessary. They could have non-linear relationship

#### **Solution 6: D**

**Explanation:** Equation for simple linear regression:  $Y=a+bx$ . Now if we increase the value of x by 1 then the value of y would be  $a+b(x+1)$  i.e. value of y will get incremented by b.

#### **Solution 7: D**

**Explanation:** Training error may increase or decrease depending on the values that are used to fit the model. If the values used to train contain more outliers gradually, then the error might just increase.

#### **Solution 8: Perfect fit for D4 and under-fit for D2.**

**Explanation:** Since is more degree 4 will be more complex (over fit the data) than the degree 3 model so it will again perfectly fit the data. In such case training error will be zero but test error may not be zero. It's highly likely that a simpler model (degree 2 polynomial) might under fit the data.