Worksheet-Set 1: Machine learning Assignment

Answer keys for questions from Q1 to Q11:

Question Number	Answer keys
Q1 :	A) Least Square Error
Q2 :	A) Linear regression is sensitive to outliers
Q3 :	B) Negative
Q4 :	B) Correlation
Q5 :	C) Low bias and high variance
Q6 :	B) Predictive modal
Q7 :	D) Regularization
Q8 :	D) SMOTE
Q9 :	A) TPR and FPR
Q10 :	B) False
Q11 :	B) Apply PCA to project high dimensional data

Answer keys for questions from Q12:

Question Number	Answer keys
Q12 :	A) We don't have to choose the learning rate. B) It becomes slow when number of features is very large. C) We need to iterate.

Q13. Explain the term regularization?
A technique used for constraining a machine learning model or calibration of machine learning model in order to make it simpler, minimize the adjusted loss function and prevent over fitting or under fitting is known as Regularization. Using Regularization, we can fit our machine learning model appropriately on a given dataset and hence reduce the errors in it.

Q14. Which particular algorithms are used for regularization?

There are two main types of regularization techniques used:

1. **Ridge Regularization (L2 Norm):**
   It modifies the over-fitted or under fitted models by adding the penalty equivalent to the sum of the squares of the magnitude of coefficients. This means that the mathematical function representing our machine learning model is minimized and coefficients are calculated. The magnitude of coefficients is squared and added. Ridge Regression performs regularization by shrinking the coefficients present.
2. **Lasso Regularization (L1 Norm):**
   It modifies the over-fitted or under-fitted models by adding the penalty equivalent to the sum of the absolute values of coefficients. Lasso regression also performs coefficient minimization, but instead of squaring the magnitudes of the coefficients, it takes the true values of coefficients. This means that the coefficient sum can also be 0, because of the presence of negative coefficients. Lasso regression line fits the model more accurately than the linear regression line.

Dropout is also a Regularization technique used in neural networks e.g. ANN, DNN, CNN or RNN to moderate the learning.

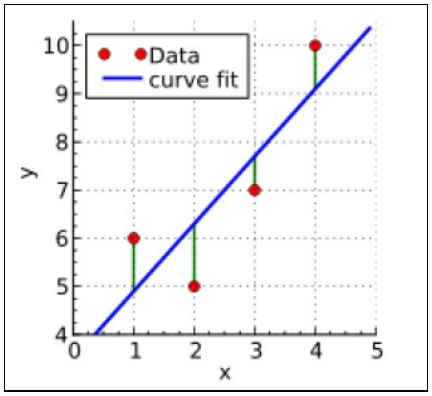
---

15. Explain the term error present in linear regression equation?

The linear regression model is stated by the equation:

$$y_i=\beta_0+\beta_1X_{1i}+\beta_2X_{2i}+\cdots+\beta_kX_{ki}+\varepsilon_i$$

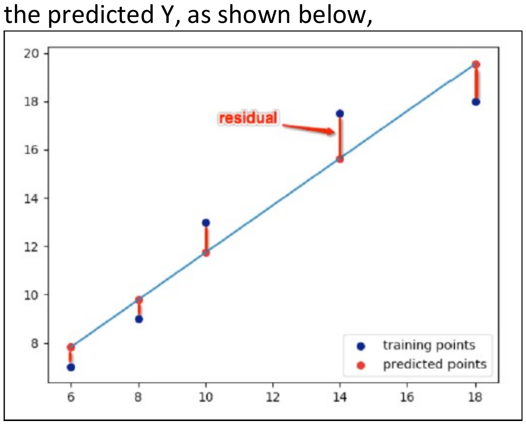
where β0 is the intercept, βi's are the slope between Y and the appropriate Xi, and ε is the error term that captures errors in measurement of Y and the effect on Y of any variables missing from the equation that would contribute to explaining variations in Y.
The residual is the observed error term (ε'hat = y– y'hat), and ε is the unobserved error term.
To explain the error further with figures & graphs, let us consider an example;



1. The red points are the observed values of x and y.
2. The blue line is the least squares line.
3. The green lines are the residuals, which is the distance between the observed values and the least squares line.

The general equation of a straight line is:
$y = mx + b$

The best fit line is obtained by minimizing the residual. Residual is the distance between the actual Y and the predicted Y, as shown below,



Mathematically, Residual is:
$r = y - (mx + b)$

$$r_i = y_i - (mx_i + b) \quad \text{(Residual for one point)}$$

$$\sum_{i=1}^{n} r_i = \sum_{i=1}^{n} (y_i - (mx_i + b)) \quad \text{(Sum of residuals)}$$

$$R(x) = \sum_{i=1}^{n} r_i^2 = \sum_{i=1}^{n} (y_i - (mx_i + b))^2 \quad \text{(Sum of squares of residuals)}$$

The R-squared (R^2) statistic provides a measure of fit. It always takes one value between 0 and 1. In simple words, For example, statistic = 0.75, it says that our model fits 75 % of the total data set. Similarly, if it is 0, it means none of the data points is being explained and a value of 1 represents 100% data explanation. Mathematically R^2 statistic is calculated as :

$$R^2 = \frac{\text{TSS} - \text{RSS}}{\text{TSS}} = 1 - \frac{\text{RSS}}{\text{TSS}}$$

RSS - Residual Sum of squares; TSS – Total Sum of squares

$$\text{RSS} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$\text{TSS} = \sum (y_i - \bar{y})^2$$