MACHINE LEARNING

1. A) Least Square Error
2. A) Linear Regression is sensitive to outliers
3. B) Negative
4. B) Correlation
5. C) Low bias and high variance
6. C) Reinforcement Learning
7. D) Regularization
8. D) SMOTE
9. C) Sensitivity and Specificity
10. B) False
11. B) Apply PCA to project high dimensional data

12. A) We don't have to choose the learning rate
    B) It becomes slow when number of features is very large

13. Regularization is a technique used to reduce the errors by fitting the function appropriately on the given training set and avoid overfitting. It is a form of regression, that constrains/ regularizes or shrinks the coefficient estimates towards zero. In other words, this technique discourages learning a more complex or flexible model, so as to avoid the risk of overfitting.

14. The commonly used regularization techniques are :
    - L1 regularization

      A regression model which uses L1 Regularization technique is called LASSO (Least Absolute Shrinkage and Selection Operator) regression. In this technique, the cost function is altered by adding the penalty term (shrinkage term), which multiplies the lambda with the absolute weight of each individual feature. Therefore, the optimization function becomes:

$$\sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2+\lambda\sum_{j=1}^{p}|\beta_j| = \mathrm{RSS}+\lambda\sum_{j=1}^{p}|\beta_j|.$$

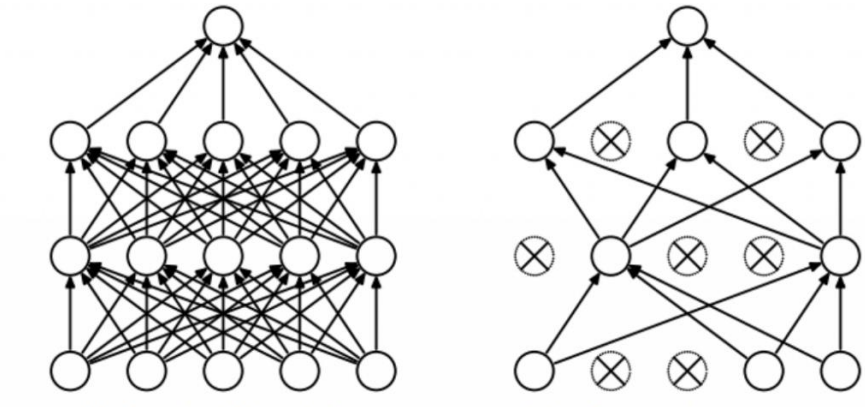
- L2 regularization

  A regression model that uses L2 regularization technique is called Ridge regression. In this technique, the cost function is altered by adding the penalty term (shrinkage term), which multiplies the lambda with the squared weight of each individual feature. Therefore, the optimization function becomes:

$$\sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2+\lambda\sum_{j=1}^{p}\beta_j^2 = \mathrm{RSS}+\lambda\sum_{j=1}^{p}\beta_j^2$$

- Dropout regularization

  Dropout is a regularization technique used in neural networks. It prevents complex co-adaptations from other neurons. In neural nets, fully connected layers are more prone to overfit on training data. Using dropout, you can drop connections with 1-p probability for each of the specified layers. Where p is called keep probability parameter and which needs to be tuned.
  With dropout, you are left with a reduced network as dropped out neurons are left out during that training iteration.



(a) Standard Neural Net (b) After applying dropout.

15. The regression equation is written as Y = a + bX +e

- Y is the value of the Dependent variable (Y), what is being predicted or explained
- a or Alpha, a constant; equals the value of Y when the value of X=0
- b or Beta, the coefficient of X; the slope of the regression line; how much Y changes for each one-unit change in X.
- X is the value of the Independent variable (X), what is predicting or explaining the value of Y
- e is the error term; the error in predicting the value of Y, given the value of X

An error term is a residual variable produced by a statistical or mathematical model, which is created when the model does not fully represent the actual relationship between the independent variables and the dependent variables. As a result of this incomplete relationship, the error term is the amount at which the equation may differ during empirical analysis.

An error term represents the margin of error within a statistical model; it refers to the sum of the deviations within the regression line, which provides an explanation for the difference between the theoretical value of the model and the actual observed results.