MACHINE LEARNING

1. Answer is (A) – Least Square Error

2. Answer is (A) - Linear regression is sensitive to outliers

3. Answer is (A) - Positive

4. Answer is (C) - Both of them

5. Answer is (C) - Low bias and high variance

6. Answer is (B) - Predictive modal

7. Answer is (D) – Regularization

8. Answer is (D) – SMOTE

9. Answer is (A) - TPR and FPR

10. Answer is (A) – True

11. Answer is (B) - Apply PCA to project high dimensional data

12. Answer is
    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.

13. Regularization is a method to constraint the model to fit our data accurately and not over fit. It can also be thought of as penalizing unnecessary complexity in our model. There are mainly 3 types of regularization techniques. They are:
    a) L1 Regularization or Lasso regularization
    b) L2 Regularization or Ridge regularization
    c) Dropout

    We use regularization whenever we suspect our model is overfitting. The biggest signs of overfitting are the poor performance of validation metrics. The validation set is part of our dataset that the model has not yet seen. As we want to detect if our model is learning just from the data, or is being heavily influenced by noise, we use the validation set which has different noise than our training set. So if our model were to over fit the training data, it would predict poorly on our validation set. During training, we also constantly measure validation metrics. If we see the validation metrics not improving significantly, or worsening, this is a telltale sign that our model is overfitting. We need to then apply regularization techniques.

    L1 Regularization

It works by adding a penalty based on the absolute value of parameters scaled by some value named lambda.

Initially our loss function was: Loss = f(preds,y)
Where y is the target output, and preds is the prediction
preds = WX + b, where W is parameters, X is input and b is bias.

With L1 regularization we add an extra term of l*|W|, where W is the weight matrix (parameters). So our loss function after L1 regularization is

Loss = f(preds, y) + l*abs(W)

L2 Regularization

L2 regularization is very similar to L1 regularization, except the penalty term is the square of the parameters scaled by some factor l (lambda)

Loss = f(preds, y) + l*$(W)^2$

Dropout

Dropout is an amazing regularization technique that works only on neural networks. The amazing idea of dropout is to randomly zero some elements of the input tensor with probability p (p is a hyperparameter). Dropout is found to work very well in practice and is simple to implement.

14. Regularization refers to a range of strategies for regularizing learning from specific characteristics in classical algorithms or neurons in neural network algorithms. It normalizes and moderates weights associated with a feature or a neuron so that algorithms aren't reliant on a small number of features or neurons to predict the outcome. This method aids in avoiding the issue of overfitting. This method reduces the computational cost of a complicated model by converting it to a simpler one to avoid overfitting.

Regularization Algorithms
Ridge regression – Its purpose is to overcome problems such as data overfitting and multicollinearity in data. When there is considerable collinearity (the existence of near-linear connections among the independent variables) among the feature variables, a typical linear or polynomial regression model will fail. Ridge Regression adjusts the variables by a modest squared bias factor. The feature variable coefficients are pulled away from this rigidity by such a squared bias factor, providing a little bit of bias into the model but considerably lowering variation.
Ridge is an excellent way to prevent overfitting.
Use regularization to solve overfitting and feature selection if you have a model with a high number of features in the dataset and want to prevent making the model too complicated.

However, the ridge has one major drawback: the final model has all N characteristics. Ridge regression decreases the two coefficients towards each other when the variables are highly linked. Lasso is torn between the two and prefers one over the other. One never knows which variable will be chosen depending on the situation. Elastic-net is a hybrid of the two that tries to shrink while still doing the sparse selection.

LASSO – It simply penalizes large coefficients, in contrast to Ridge Regression. When the hyperparameter is big enough, Lasso has the effect of driving certain coefficient estimations to be absolutely zero. As a result, Lasso conducts variable selection, resulting in models that are significantly easier to read than Ridge Regression models. In a nutshell, it's about lowering variability and increasing the accuracy of linear regression models. If we have a large number of features, LASSO works effectively for feature selection. It reduces coefficients to zero and if a set of predictors is highly associated, lasso selects one and reduces the others to zero.

15. 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. The error term is also known as the residual, disturbance, or remainder term, and is variously represented in models by the letters e, ε, or u.
-An error term appears in a statistical model, like a regression model, to indicate the uncertainty in the model.
-The error term is a residual variable that accounts for a lack of perfect goodness of fit.
-Heteroskedastic refers to a condition in which the variance of the residual term, or error term, in a regression model varies widely.