Machine Learning

1. High R-squared value for train-set and Low R-squared value for test-set.
2. Decision trees are highly prone to overfitting
3. Random Forest
4. Precision
5. Model B
6. Lasso and Ridge
7. Random Forest and Decision Tree
8. Pruning and Restricting the max depth of the tree
9. We initialize the probabilities of the distribution as 1/n, where n is the number of data-points and A tree in the ensemble focuses more on the data points on which the previous tree was not performing well
10. The adjusted R-squared compensates for the addition of variables and only increases if the new predictor enhances the model above what would be obtained by probability. Conversely, it will decrease when a predictor improves the model less than what is predicted by chance.
11. **Ridge Regression** : In ridge regression, the cost function is altered by adding a penalty equivalent to square of the magnitude of the coefficients. So ridge regression puts constraint on the coefficients (w). The penalty term (lambda) regularizes the coefficients such that if the coefficients take large values the optimization function is penalized. So, ridge regression shrinks the coefficients and it helps to reduce the model complexity and multi-collinearity

    Lasso Regression : In Ridge regression,the only difference is instead of taking the square of the coefficients, magnitudes are taken into account. This type of regularization (L1) can lead to zero coefficients i.e. some of the features are completely neglected for the evaluation of output. So Lasso regression not only helps in reducing over-fitting but it can help us in feature selection.

12. Variance inflation factor (VIF) is a measure of the amount of multicollinearity in a set of multiple regression variables. There is no suitable value of a VIF for a feature to be included in a regression modeling, however VIF exceeding 5 or 10 indicates high multicollinearity between this independent variable and the others.

13. Scaling is necessary to ensure that the gradient descent moves smoothly towards the minima and that the steps for gradient descent are updated at the same rate for all the features, we scale the data before feeding it to the model. This also ensures that there is no biasedness present by ensuring equal weightage to every data points.

14. The metrics used to check the goodness of fit in Linear Regression are Mean Squared Error (MSE). Root Mean Squared Error (RMSE). Mean Absolute Error (MAE).

15. Sensitivity= 1000/1000+50= 51
    Specificity= 1200/1200+250= 251
    Precision= 1000/1000+250= 251
    Recall= 1000/1000+50= 51
    Accuracy= 1000+1200/1000+50+250+1200= 1502.2

Statistics

1. The outcome of exam
2. Discrete
3. PDF
4. Mean
5. empirical mean
6. Standard Deviation
7. 0 and 1
8. Bootstrap
9. Summarized.
10. Histograms are preferred to determine the underlying probability distribution of a data. Box plots on the other hand are more useful when comparing between several data sets. They are less detailed than histograms and take up less space.
11.
12. Statistical significance is often calculated with statistical hypothesis testing, which tests the validity of a hypothesis by figuring out the probability that your results have happened by chance.
13. Any distribution of money or value will be non--Gaussian. For example: distributions of income; distributions of house prices; distributions of bets placed on a sporting event. These distributions cannot have negative values and will usually have extended right hand tails.
14. The median better represents the central tendency for the skewed distribution. Income is the classic example of when to use the median instead of the mean because its distribution tends to be skewed.
15. The likelihood function (often simply called the likelihood) is the joint probability of the observed data viewed as a function of the parameters of the chosen statistical model.