Worksheet 6

MACHINE LEARNING

1	1. In which of the following you can say that the model is overfitting? A) High R-squared value for train-set and High R-squared value for test-set. B) Low R-squared value for train-set and High R-squared value for test-set. C) High R-squared value for train-set and Low R-squared value for test-set. D) None of the above
Ans	C) High R-squared value for train-set and Low R-squared value for test-set.
2	Which among the following is a disadvantage of decision trees? A) Decision trees are prone to outliers. B) Decision trees are highly prone to overfitting. C) Decision trees are not easy to interpret D) None of the above.
Ans	B) Decision trees are highly prone to over fitting
3	Which of the following is an ensemble technique? A) SVM B) Logistic Regression C) Random Forest D) Decision tree
Ans	C) Random Forest
4	Suppose you are building a classification model for detection of a fatal disease where detection of the disease is most important. In this case which of the following metrics you would focus on? A) Accuracy B) Sensitivity C) Precision D) None of the above.
Ans	B) Sensitivity
5	The value of AUC (Area under Curve) value for ROC curve of model A is 0.70 and of model B is 0.85. Which of these two models is doing better job in classification? A) Model A B) Model B C) both are performing equal D) Data Insufficient In
Ans	B) Model B
6	Which of the following are the regularization technique in Linear Regression?? A) Ridge B) R-squared C) MSE D) Lasso
Ans	A) Ridge D) Lasso
7	Which of the following is not an example of boosting technique? A) Adaboost B) Decision Tree C) Random Forest D) Xgboost.
Ans	B) Decision Tree C) Random Forest
8	Which of the techniques are used for regularization of Decision Trees? A) Pruning B) L2 regularization C) Restricting the max depth of the tree D) All of the above
Ans	A) Pruning and C) Restricting the max depth of the tree
9	Which of the following statements is true regarding the Adaboost technique? A) We initialize the probabilities of the distribution as 1/n, where n is the number of data-points B) A tree in the ensemble focuses more on the data points on which the

	previous tree was not performing well C) It is example of bagging technique D) None of the above
Ans	B) A tree in the ensemble focuses more on the data points on which the previous tree was not performing well
10	Explain how does the adjusted R-squared penalize the presence of unnecessary predictors in the model?
	The adjusted R-squared is a modified version of the R-squared metric used in linear regression models. The adjusted R-squared penalizes the presence of unnecessary predictors in the model by adjusting the R-squared value based on the number of predictors used. In a linear regression model, the R-squared value represents the proportion of the variance in the dependent variable that is explained by the independent variables. However, adding more predictors to the model does not necessarily improve the R-squared value. In fact, adding unnecessary predictors can sometimes even decrease the model's predictive power by introducing noise into the model. The adjusted R-squared adjusts the R-squared value based on the number of predictors used in the model. Specifically, the adjusted R-squared is calculated as follows: Adjusted R-squared = 1 - [(1 - R-squared) * (n - 1) / (n - k - 1)] where n is the number of observations in the dataset, and k is the number of predictors in the model. The adjusted R-squared penalizes the presence of unnecessary predictors in the model by subtracting a penalty term from the R-squared value based on the number of predictors used. The penalty term is proportional to the number of predictors in the model and is larger when more predictors are used. Therefore, the adjusted R-squared provides a more accurate measure of the model's predictive power by taking into account the number of predictors used and penalizing the presence of unnecessary predictors in the model. This helps to prevent overfitting and ensures that the model is not including predictors that do not add significant value to the model.
11	Differentiate between Ridge and Lasso Regression.
	<table><tr><th>Ridge Regression</th><th>Lasso Regression</th></tr><tr><td>Ridge regression, also known as L2 regularization, adds a penalty term to the sum of squared errors (SSE) objective function in linear regression. The</td><td>Lasso regression, also known as L1 regularization, adds a penalty term to the SSE objective function that is proportional to the absolute magnitude</td></tr></table>

	penalty term is proportional to the squared magnitude of the coefficients, and it shrinks the coefficients towards zero but does not set them exactly to zero. As a result, Ridge regression can reduce the variance of the model without significantly affecting its bias.	of the coefficients. Lasso regression not only shrinks the coefficients towards zero, but it also can set some of them to exactly zero. This means that Lasso regression can be used for feature selection, where irrelevant predictors are eliminated from the model.
	Reduces the impact of multicollinearity among the predictors.	Can be used for feature selection by eliminating irrelevant predictors.
	Can handle cases where the number of predictors is larger than the number of observations.	Can be more interpretable than Ridge regression because it can reduce the number of predictors to a few important ones.
	Works well when all the predictors are important.	Works well when some of the predictors are less important or redundant

12	What is VIF? What is the suitable value of a VIF for a feature to be included in a regression modelling?
	VIF stands for Variance Inflation Factor, which is a measure of the degree of multicollinearity among the predictor variables in a regression model. It quantifies the extent to which the variance of the estimated regression coefficient for a predictor is inflated due to the correlation with the other predictors in the model. The VIF for a given predictor is calculated as 1/(1-R^2), where R^2 is the coefficient of determination obtained from a regression model that uses all other predictors to predict the given predictor. The higher the VIF, the higher the degree of multicollinearity. A common rule of thumb is that a VIF value of 1 indicates no correlation among predictors, while a value of 5 or above indicates a high degree of correlation. Therefore, a suitable value for VIF for a feature to be included in a regression model is typically less than 5. However, the exact threshold value for VIF can vary depending on the context and the specific goals of the analysis. It is generally recommended to assess the VIF values for all the predictors in the model and remove any predictors with VIF values above the threshold to address the issue of multicollinearity. By reducing multicollinearity, we can improve the stability and accuracy of the regression coefficients and the predictions made by the model.
13	Why do we need to scale the data before feeding it to the train the model
	We need to scale the data before feeding it to train the model because many machine learning algorithms work better or require the data to be in a particular range or scale. There are several reasons for this:

	Different units: The features in the dataset may be measured in different units or scales. For example, one feature may be measured in dollars, while another may be measured in kilograms. This difference in units can result in some features dominating others in the model. Gradient descent optimization: Many machine learning algorithms use gradient descent optimization, which works better when the features are in a similar range. If the features have very different ranges, the gradient descent algorithm may take longer to converge or may get stuck in local minima. Distance-based algorithms: Some algorithms like K-Nearest Neighbors (KNN) and Support Vector Machines (SVM) are distance-based algorithms, and they are sensitive to the scale of the features. Features with larger scales will have a larger effect on the distance calculations, leading to biased results. Scaling the data involves transforming the features to have a similar range, typically between 0 and 1 or with a mean of 0 and standard deviation of 1. Common methods for scaling include Min-Max scaling, Standardization, and Normalization. Scaling the data ensures that all the features are on the same scale and contribute equally to the model, resulting in better performance and more accurate predictions.
14	What are the different metrics which are used to check the goodness of fit in linear regression?
	Linear regression is a statistical method used to analyze the relationship between two continuous variables. There are several metrics that can be used to evaluate the goodness of fit in linear regression, including: R-squared (R^2) : It is the proportion of the variation in the dependent variable (y) that is explained by the independent variables (x). R-squared values range from 0 to 1, with higher values indicating a better fit of the model. Mean Squared Error (MSE): It measures the average of the squared differences between the predicted and actual values. The lower the MSE, the better the fit of the model. Root Mean Squared Error (RMSE): It is the square root of the MSE and measures the average magnitude of the error in the units of the dependent variable. The lower the RMSE, the better the fit of the model. Mean Absolute Error (MAE): It measures the average absolute difference between the predicted and actual values. The lower the MAE, the better the fit of the model. Residual Standard Error (RSE): It is similar to RMSE, but normalized by the degrees of freedom of the residuals. It measures the variability of the residuals and the goodness of fit of the model.

	Adjusted R-squared: It is a modified version of R-squared that takes into account the number of independent variables in the model. It penalizes the inclusion of unnecessary variables in the model and helps to prevent overfitting. Overall, a combination of these metrics should be used to evaluate the goodness of fit in linear regression and select the best model for predicting the dependent variable.
15	From the following confusion matrix calculate sensitivity, specificity, precision, recall and accuracy
	From the given confusion matrix, we can calculate the following metrics: True Positive (TP) = 1000 False Positive (FP) = 250 True Negative (TN) = 1200 False Negative (FN) = 50 Sensitivity = TP / (TP + FN) = 1000 / (1000 + 50) = 0.952 Specificity = TN / (TN + FP) = 1200 / (1200 + 250) = 0.82 Precision = TP / (TP + FP) = 1000 / (1000 + 250) = 0.8 Recall = Sensitivity Accuracy = (TP + TN) / (TP + TN + FP + FN) = (1000 + 1200) / (1000 + 1200 + 50 + 250) = 0.9 Therefore, the sensitivity is 0.8, specificity is 0.96, precision is 0.95, recall is 0.8, and accuracy is 0.9.