# ASSIGNMENT

# Of D4 and d5 questions

1. d) Collinearity Dimensionality reduction can help reduce collinearity by reducing the number of correlated features in a dataset.
2. b) Random Forest Random Forest is based on the idea of bagging (Bootstrap Aggregating).
3. c) Decision Tree are prone to overfit One of the disadvantages of decision trees is that they are prone to overfitting.
4. c) Training data Machine learning algorithms build a model based on training data.
5. c) Anomaly detection Anomaly detection techniques help in detecting outliers in data.
6. a) Support Vector Support Vector is not a numerical function in the context of function representations of machine learning algorithms.
7. d) Both a and b Analysis of ML algorithms can involve both statistical learning theory and computational learning theory.
8. c) Both a and b The difficulties with the k-nearest neighbor algorithm include the curse of dimensionality and the need to calculate the distance of the test case for all training cases.
9. a) 1 Radial Basis Function (RBF) neural networks typically have one layer.
11. c) Neither feature nor number of groups is known

Unsupervised learning is when neither the features nor the number of groups are explicitly known.

1. b) SVG

SVG is not a machine learning algorithm.

1. b) Underfitting

Underfitting occurs when the model fails to decipher the underlying trend in the input data.

1. a) Reinforcement

learning Real-Time decisions, Game AI, Learning Tasks, Skill acquisition, and Robot Navigation are applications of reinforcement learning.

1. b) Mean squared error

Mean squared error is the average squared difference between the classifier's predicted output and the actual output.

1. a) Linear, binary

Logistic regression is a linear regression technique used to model data with a binary outcome.

1. A. supervised learning

Classifying reviews of a new Netflix series is an example of supervised learning, where the labels (positive, negative, neutral) are provided for training.

1. C. both a and b

Euclidean distance and Manhattan distance are both powerful distance metrics used in geometric models.

1. B. removing columns which have high variance in data

Removing columns with high variance can help in reducing the dimensions of a dataset.

1. C. input attribute.

Both supervised learning and unsupervised clustering require input attributes.

1. A) -(6/10 log(6/10) + 4/10 log(4/10))

This formula represents the entropy of the target variable.

1. A) weights are regularized with the l1 norm

Lasso regression is interpreted as least-squares linear regression with weights regularized using the l1 norm.

1. B) increase by 5 pounds

According to the linear regression equation, for each one-inch increase in height, the weight should increase by 5 pounds.

1. D) Minimize the squared distance from the points

The least squares regression line minimizes the squared distance from the points.

1. B) As the value of one attribute increases, the value of the second attribute also increases

A correlation coefficient of 0.85 indicates a positive linear relationship between the two attributes.

1. B) Convolutional Neural Network

Convolutional Neural Networks (CNNs) are well-suited for image identification problems like recognizing objects in photos.

1. a) PCA Principal Component Analysis (PCA) is an unsupervised learning technique, not a supervised learning technique.