Postlab 5

13/10/23

9237
Ryan V
Batch :- D

1. SVM can be used for both classification as well as regression. It helps to find the hyperplane that maximally separates different classes of data while maintaining the largest margin between the classes.

2. In the context of SVMs, Convex Hull is the outer boundary formed by the support vector and is critical in defining the margin and SVMs decision boundary. It represents the region in which SVM finds the optimal hyperplane for classification.

3. Hard Margin :- It seeks to find hyperplane that perfectly separates 2 classes of data points without any missclassification.

Soft Margin :- This allows missclassification to a certain extent.

4. Hinge Loss :-

This loss is used in ML, mainly in SVM and binary classification tasks.
It is designed to quantify the error.

$$Loss(y, f(x)) = \max(0, 1 - y * f(x))$$

5 "Kernel Trick" is a fundamental concept in ML. It is used to implicitly map data from a lower-dimensional space to a higher-dimensional space without explicitly computing the transformation.

6 Explain about SVM regression

→ It is used for regressions tasks. While traditional SVMs are designed for classification tasks, it also helps to predicting continuous numeric values.

13/10/23

Postlab 6

9237
Ryan . V.

1. Similarity - based clustering is a technique in unsupervised learning algo. It uses similarity measures to compare data points and groups points into clusters based on their dissimilarity or similarity.

2. Significance testing in clustering is crucial for validating and ensuring the reliability of the obtained clusters, aiding in their interpretation and making informed decisions about clustering methods & parameters.

3 i) Customer segmentation.
ii) Image compression
iii) Healthcare

4.

	Hard Clustering	Soft Clustering
i)	In this each data point belongs exclusively to one point.	Some points may belong to multiple clusters.
ii)	Each point is assigned to a single cluster.	They are associated with a set of clusters.
iii)	Kmeans, Hierarchical	Fuzzy c-means, GMM.

5 It is difficult to determine the optimal no. of clusters.
The algos are sometimes sensitive to order of data.
Results may change based on how data is arranged.

6 It's difficult as from data you won't come to know how many clusters are there and if data is clusterable or not.
Clusters may vary in density and may not be of equal size.

7

Partition Clustering	Hierarchical Clustering
i) It aims to divide the dataset into a set of non-overlapping clusterings, where each data point belongs exclusively to one cluster.	It constructs a tree like hierarchy of clusters where data points can belong to multiple clusters at diff levels.
ii) Need To specify no of clusters.	No need To specify
Kmeans, Kmedoids	Agglomerative & divisive clustering.

13/10/23

Postlab 7

Ryan .V.
9237.

1 Weak learners are models that perform slightly better than random guessing or chance on a classification or regression task.
The models are characterized by their limited predictive power when used individu

2 The key idea behind a Random Forest is that by combining multiple trees to add randomness.
It reduces overfitting caused by decision tree.

3 Bagging involves multiple base models on different random subsets of the training data, created through bootstrapping.
Bagging reduces variance & helps to preven overfitting.

Boosting uses multiple base models & sequentially trains them.
Each base model is trained to correct errors.
Boosting is effectime at reducing bias.

4 Stacking is an ensemble learning Techniqu that combines the predictions from multipl base models.
It leverages strength from various model and combines them.

5 It combines multiple algos in a hierarchical fashion to make predictions. It is especially good at dealing with complex or noisy datasets.

6 Meta-learning focuses on training models how to learn.
The idea here is to leverage the knowledge gained from previous tasks to facilitate faster and more accurate learning on new, unseen tasks.

13/10/23

Postlab 8

9237
Ryan . V.

1.) PCA is a dimensionality reduction technique
It's main purpose is to transform high-dimensional data into a lower one.
It also tries to preserve as much of the data's variability as possible.

2) It's the process of reducing the no. of variables in the dataset.
This is to reduce the complexity of model.
It also tries to capture the variability of the ~~ng~~ dataset
This makes it more manageable for analysi

3) The curse of dimensionality is a major complication in ML.
It refers to datasets having many featur
It increases computational complexities, increases overfitting and difficulty in visualization.

4) Hyperparameter Tuning is a method of finding the optimal parameters to ensure the best-fit model.
Some common methods are:-
i) Grid Search.
ii) Random Search.
iii) Bayesian Optimisation.