Page No.
Date

NAME: AFNAN ATTAR PRN:- F19112003 CLASS: BE COMPIT
SUBJECT:- ML EXPERIMENT NO.:- A4

Q1) Compare the Mini-batch Gradient Descent, Stochastic Gradient Descent, Batch Gradient Descent.

Ans Let us compare Mini-batch, Stochastic and Batch Gradient Descent.

Comparision	Mini-batch	Stochastic	Batch
1. Defination	Split dataset and compute gradient descent for each batch.	Computes Gradient on the basis of one randomly choosen partition of dataset.	Iteratively finds minimum of any differentiable function.
2. Formula Used	$w_{i+1} = w_i - a \cdot \nabla_{w_i} J(x^{i:i+b}, y^{i:i+b}; w_i)$	$w_{i+1} = w_i - a \cdot \nabla_{w_i} J(x^i, y^i; w_i)$	$w_{i+1} = w_i - a \cdot \nabla_{w_i} J(w_i)$
3. Speed	~~Time~~ Fastest	Faster	Time-Consuming
4. Computation	~~Expensive~~ Lightest	Lighter	Expensive
5. Practically Used	Often	Sometimes	Rare

Q2) Explain how does the gradient descent work in linear regression.

Ans: In linear regression, the model targets to get best-fit regression line to predict the values of y based on the given input value (x).

2. The model calculates the cost function which measures the Root Mean square error between predicted value (pred) and true value (y).
3. To minimize cost function, the model needs to have best value of θ_1 and θ_2.
4. To do this we use gradient descent, it steps down the cost function in the direction of steepest descent.

$$\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$$

$$\therefore \theta_j = \theta_j - \frac{\alpha}{m} \sum_{i=1}^{m} [(h_\theta(x_i) - y_i) x_i]$$

Q3) Does Gradient Descent always converge to an optimum?

Ans 1. No, gradient descent does not always converge to an optimum.
2. If the function is convex and the step size is not too large, it will converge to within some tolerance.
3. If the function is non-convex, gradient descent will have a hard time converging to a local minima.
4. As gradient descent takes blind steps it can overshoot and go past local minima.

NAME: AFNAN ATTAR PRN:- F19112003 CLASS:- BE COMP II
SUBJECT: ML EXPERIMENT No.: 05 (A)

Q1) Is feature scaling required for the KNN Algorithm? Explain with proper justification.

Ans 1. Yes, feature scaling is required to get the better performance of the KNN algorithm.

2. For example, imagine a dataset having n number of instances and N number of features. There is one feature that has values between 0 and 1, meanwhile another feature airs between -999 to 999, this would cause biasing.

Q2) How can you relate KNN algorithm to the Bias-Variance tradeoff?

Ans 1. The bias will be 0 when k=1, however, when it comes to new data (in test set), it has higher chance to be an error, which causes high variance.

2. When we increase K, the training error will increase (increase bias), but the test error may decrease at the same time (decrease variance).

Q3) The KNN algorithm does more computation on test time rather than train time.

Ans 1. Yes, KNN spends more computation on test time rather than train time.

2. The idea of KNN algorithm is to find a k-long list of samples that are close to a sample we want to classify.

3. Therefore, the training phase is basically storing

a training set, whereas while the prediction stage the algorithm looks for k-neighbours using that stored data.

NAME: AFNAN ATTAR PRN:- F19112003 SUBJECT:- ML
CLASS:- BE COMP II EXPERIMENT No.:- A6

Q1) Compare Hierarchial Clustering and K-Means Clustering

Ans

Hierarchial Clustering	K-Means Clustering
1. Hierarchical methods can be either divisive or agglomerative.	Assigns records to each cluster to find mutually exclusive cluster.
2. One can stop at any number of clusters.	It requires advance knowledge of K, i.e number of clusters.
3. Sequentially combines 'n' clusters until one remains.	One uses mean or median as a cluster centre to represent each cluster.
4. Used when target is to arrange clusters into a naturally hierarchy.	Methods used are normally less computationally intensive and are used for large dataset.
5. Results are reproducible	Results are not reproducible

Q2) What are some stopping criteria for K-Means clustering?

Ans Here are some stopping criterias for K-Means:-

1. Convergence. (No further changes).
2. Maximum no. of iterations reached.
3. Variance did not improve by atleast x.
4. Variance did not improve by atleast x* intial variance.

Q3) How would you Pre-Process the data for K-Means?

Page No.
Date

Ans Stages of Data preprocessing for K-Means clustering:-

1. Data cleaning:-
   - Removing duplicates, irrelevant observation and errors.
   - Handling inconsistent data, outliers & noise.
2. Handling Missing data.
3. Data Integration.
4. Data Transformation:-
   - Feature Construction
   - Handling skewness
   - Data Scaling.
5. Data Reduction:-
   - Removing highly correlated variables.
   - Feature Selection.
   - PCA.