COMPUTER ENGINEERING DEPARTMENT

ASSIGNMENT NO-02

SUB: Machine Learning

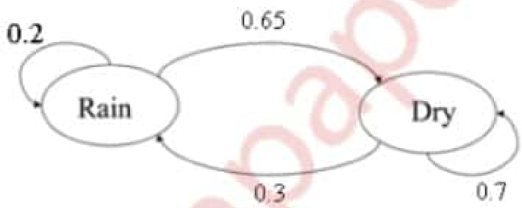
COURSE: T.E. **Year: 2020-2021** **Semester: VI**
DEPT: Computer Engineering
SUBJECT CODE: CSDLO6021 **SUBMISSION DATE: 03/05/2021**

===

Name: Amey Thakur **Roll No.: 50**

Class: TE Comps B-50 **ID: TU3F1819127**

Assignment No 2

Sr. No.	Question	CO mapping
1	Consider Markov chain model for 'Rain' and 'Dry' is shown in following figure.  Two states: 'Rain' and 'Dry'. Transition probabilities: P('Rain'|'Rain') = 0.2, P('Dry'|'Rain') = 0.65, P('Rain'|'Dry') = 0.3, P('Dry'|'Dry') = 0.7, Initial probabilities: say P('Rain') = 0.4, P('Dry') = 0.6. Calculate a probability of a sequence of states {'Dry', 'Rain', 'Rain', 'Dry'}.	CO5
2	Discuss the following terms Initial hypothesis, Expectation step, Maximization step w.r.t EM algorithm and discuss how initial hypothesis converges to an optimal solution with suitable example.	CO5
3	Discuss the structure of RBFN and how it can be used to solve non-linearly separable pattern?	CO5

4	Describe the two methods for dimensionality reduction and how Principal Component Analysis is carried out to reduce the dimensionality of data.	CO6
5	Why Dimensionality reduction is a very important step in Machine Learning? Use Principle component analysis (PCA) to arrive at the transformed matrix for the given matrix A. A^T $\begin{bmatrix} 2 & 1 & 0 & -1 \\ 4 & 3 & 1 & 0.5 \end{bmatrix}$	CO6
6	Describe the down Hill Simplex method. Why it is called the Derivative-Free method?	CO3
7	Minimize $f(x_1, x_2) = 4x_1 - 2x_2 + 2x_1^{2} + 2x_1x_2 + x_2^2$ With starting point $X_1 = \{0\ 0\}$ using the steepest decent method. (Perform two iteration)	CO3
8	Find the Singular value decomposition of $A = [2\ 2\ 1\ -1]$ and also List some advantages of derivative-based optimization techniques	CO3

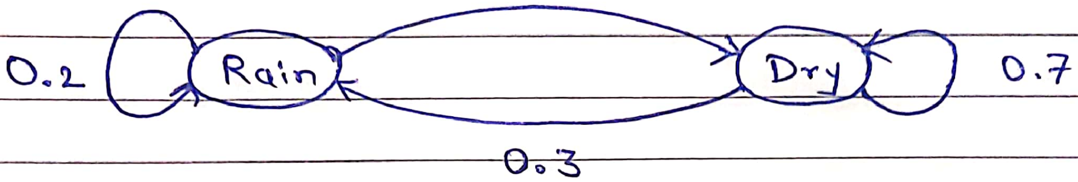
Amey

AMEY B 50 Amey

Page No.:

Date:

Q1. Consider Markov chain model for 'Rain' and 'Dry' is shown in following figure.



Two states : Rain and Dry, transition probability.

$P(\text{Rain} \mid \text{Rain}) = 0.2$

$P(\text{Dry} \mid \text{Rain}) = 0.65$

$P(\text{Rain} \mid \text{Dry}) = 0.3$

$P(\text{Dry} \mid \text{Dry}) = 0.7$

Initial Probabilities : $P(\text{Rain}) = 0.4$

$P(\text{Dry}) = 0.6$

Calculate : Probability of sequence {Dry, Rain, Rain, Dry}

Ans :

By Markov chain property,

$$P(S_{i1}, S_{i2}, \ldots, S_{ik}) = P(S_{i1}, S_{i2}, \ldots S_{ik-1}) \cdot P(S_{i1}, S_{i2}, \ldots, S_{ik})$$
$$= P(S_{ik} \mid S_{ik-1}) \cdot P(S_{i1}, S_{i2}, \ldots S_{ik-1})$$
$$= P(S_{ik} \mid S_{ik-1}) \cdot P(S_{ik-1} \mid S_{ik-2}) \ldots \ldots$$
$$\ldots \ldots P(S_{i2} \mid S_{i1}) \cdot P(S_{i1})$$

$P(\{\text{Dry}, \text{Rain}, \text{Rain}, \text{Dry}\})$

$= P(\text{Dry} \mid \text{Rain}) \cdot P(\text{Rain} \mid \text{Rain}) \cdot P(\text{Rain} \mid \text{Dry}) \cdot P(\text{Dry})$

$= 0.65 \times 0.2 \times 0.3 \times 0.6$

$= \underline{0.0234}$

AMEY B 50 Amey

Page No.:
Date:

Q.2. Discuss the following terms:
Initial hypothesis, Expectation step, Maximization step w.r.t. EM algorithm and discuss how initial hypothesis converges to an optimal solution with example.

Ans:

Initial Hypothesis

- A set of initial values of the parameters are considered, a set of incomplete observed data is given to a system with assumption that the observed data comes from specific model

Exception Step (E-step)

- In this step, we observed data in order to estimate or guess the values of missing or incomplete data it is basically used to update variables.

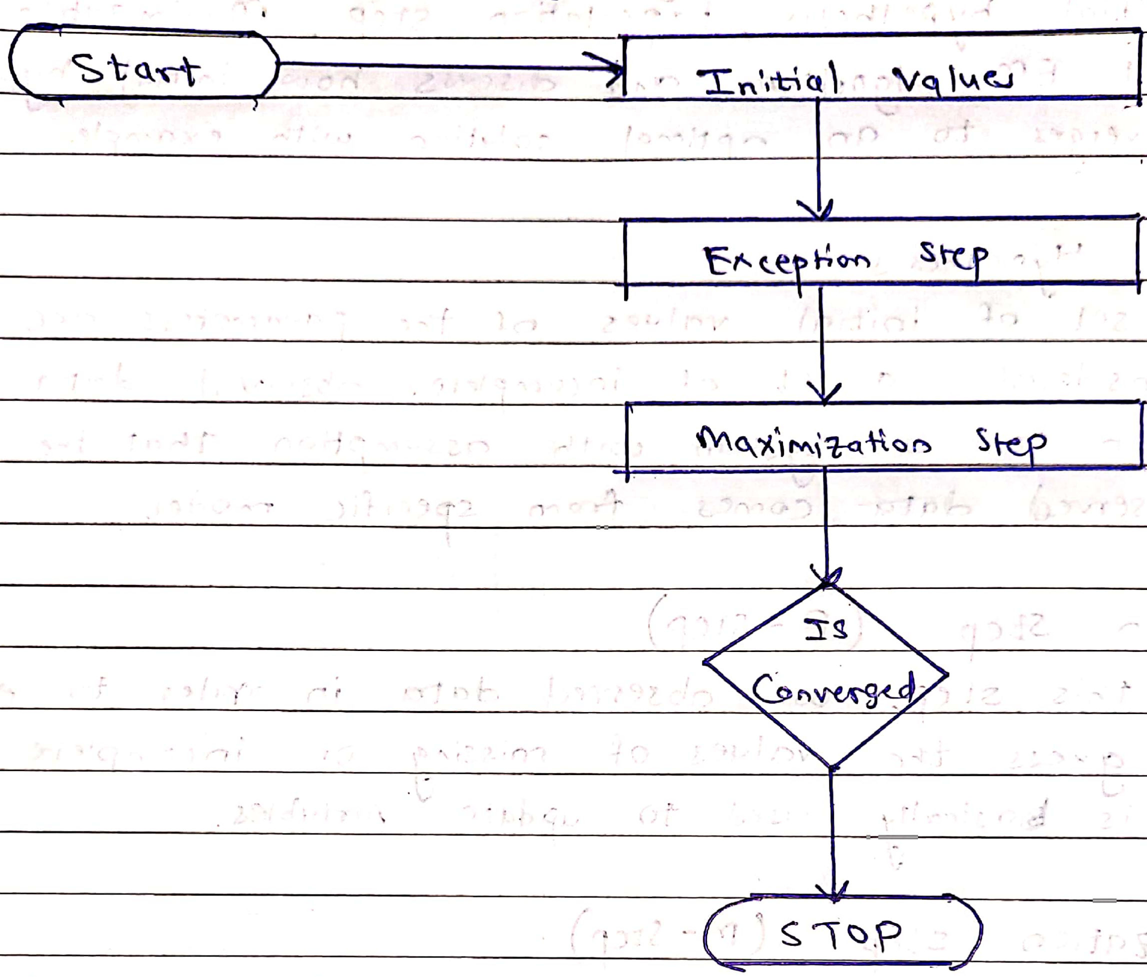
Maximization step (M-step)

- In this step, we use complete data generated in precedence exception in order to update values or performance. It is basically used to update hypothesis.

AMEY B 50 Amey

Page No.:
Date:

Flowchart of EM Algorithm



Example:

				Coin A	Coin B
		0.45×A	0.55×B	2.2H, 2.2T	2.8H, 2.8T
1	HTTTHHTHTH	0.80×A	0.20×B	7.2H, 0.8T	1.8H, 0.2T
2	HHHHTHHHHH	0.73×A	0.27×B	5.9H, 1.5T	2.1H, 0.5T
3	HTHHHHHTHH	0.35×A	0.65×B	1.4H, 2.1T	2.6H, 3.9T
4	HTHTTTHHTT	0.65×A	0.35×B	4.5H, 1.9T	2.5H, 1.1T
5	THHHTHHHTH			21.2H, 8.5T	11.8H, 8.5T

2

$\theta_A = 0.60$

$\theta_B = 0.50$

1

3 M-step

$$\theta_A = \frac{21.2}{21.2+8.5} = 0.713$$

$$\theta_B = \frac{11.8}{11.8+8.5} = 0.581$$

4

1A = 0.80

1B = 0.52

AMEY B 50 Amey

Page No.:
Date:

Q3. Discuss the structure of RBFN and how it can be used to solve non-linearly separable pattern?

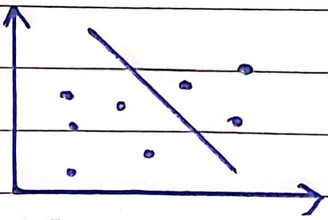
Ans:

Kernel:

- A kernel is a similarity function.
- Some algorithms use a set of mathematical functions that are defined as the kernel
- The function of the kernel is to take data as input and transform it into required form
- It is a function provided to a machine learning algorithm
- Example: Linear, Non-linear, Polynomial, Radial basis function.

Classifying non-linearly separable data:

1. To predict if a dog is a particular breed we load millions of dog's information like type, height, skin color.
2. In machine learning language these properties are referred as features.
3. A single entry of these lists of features is a data instance which the collection of everything is training data which forms the basis of your reduction.
4. SVM

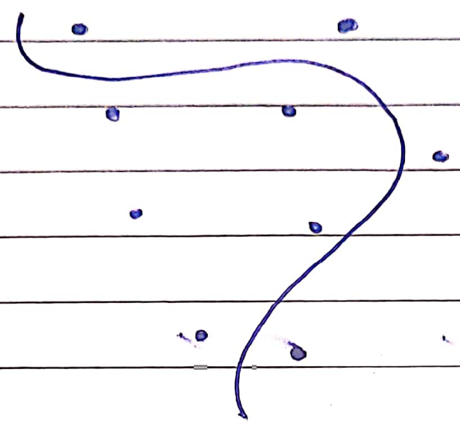


5. The hyperplane of a two dimensional space is a line dividing the red and blue balls.
6. Data (all breeds of dog) + features (skin color, health) → learning algorithm

AMEY B 50 Amey

Page No.:
Date:

(7) If we want to solve following example in linear manner, it is not possible to separate by straight line.



(a) In machine learning "kernel" is usually used to refer to the kernel trick, a method of using a linear classifier to solve a non-linear problem.

(b) It entails transforming linearly inseparable data like linearly separable ones.

(c) The kernels' function is what is applied on each data instance to map the original non-linear observation into a higher dimensional space in which they become separable.

(d) Using dog example, instead of defining a slew of features you define a single kernel function to compute similarity between breeds ~~of~~ dog.

(e) You provide this to kernel, together with data and labels to the learning algorithm and outcomes a classifier.

(f) Mathematical Function : $k(x, y) = < f(x) \cdot f(y) >$

- k = kernel function
- $x \cdot y$ are n dimensional inputs
- f is a map from n dimensional to m-dimensional space (usually $m > n$).

AMEY B 50 Amey

Page No.:
Date:

Q4. Describe the two methods for dimensionality reduction and how principal component analysis is carried out to reduce the dimensionality of data.

Ans:

- Dimensionality reduction is the process of reducing the number of random variables under consideration, by obtaining a set of principle variables.
- The methods used for dimensionality reduction are:

① Linear Discriminant Analysis (LDA)

- LDA is a type of linear combination, a mathematical process, using various data items and applying a function to that site to separately analyze multiple classes of objects or items.

Steps for performing LDA

(a) Compute the d-dimensional mean vector for the different classes from the dataset.

(b) Compute the scatter matrix.

(c) Sort the edge vector by decreasing eigen values and choose k eigen vector with the largest eigen value to form a d * k dimensional matrix w.

(d) Use d * k eigen vector matrix to transform the sample into the new subject

$$Y = X * w.$$

② Principal Component Analysis (PCA)

- PCA aims to find the directions of maximum variance in high dimensional data and projects onto a new subspace with equal or fewer dimensions than the original one.

AMEY B 50 Amey

Page No.:
Date:

Steps for performing PCA.

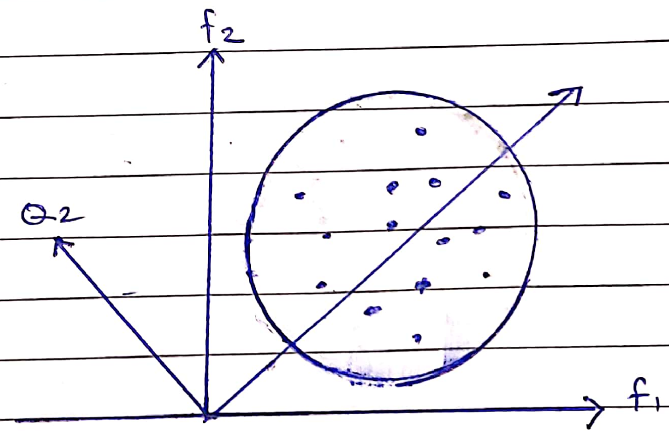
(a) Standardize the d-dimensional datasets.
(b) Construct the covariance matrix.
(c) Decompose the covariance matrix into the eigen vector with its eigen values.
(d) Select k eigen vectors which corresponds to the k largest values where k is the dimensionality of the new feature subspace ($k \leq d$).
(e) Construct a projection matrix w from the top k eigen vectors.
(f) Transform d-dimensional input dataset x using the projection matrix w to obtain new k dimensional feature subspace.

PCA for dimensional reduction

- It works on a condition that while the data in a higher dimensional space is mapped to data in a lower dimensional space, the variance of the data in lower dimensional space should be maximum.

It involves following steps:

- Construct the covariance matrix of the data.
- Compute the eigen vector of this matrix.
- Eigen vectors corresponding to the largest eigen values



are used to reconstruct a large fraction of variance of original data.

- Hence we are left with a lesser number of eigen vector and there might have been some data lost in process. But the most important variances should be retained by remaining vectors.

AMEY B 50 Amey

Page No.:

Date:

Q5. Why dimensionality reduction is a very important step in machine learning & Use Principal Component Analysis (PCA) to arrive at the transformed matrix for the given matrix A.

$$A^T = \begin{bmatrix} 2 & 1 & 0 & -1 \\ 4 & 3 & 1 & 0.5 \end{bmatrix}$$

Ans:

- High dimensional data means high computational cost to perform learning.
- It often leads to over-fitting when learning a model, which means model will perform well on training data but poorly on test data.
- Distances between a nearest and farthest data point can become equidistant in high dimensions, that can hamper the accuracy of some distant based analysis tools.

Dimensionality reduction helps with these problems, while trying to preserve most of the relevant information in the data needed to learn accurate, predictive models.

- It reduces time and storage space required.
- It helps to remove multi-collinearity which improves the interpretation of parameters of machine learning model.
- It becomes easier to visualize data when reduced to very low dimensions such as 2D or 3D.
- It avoids the cause of dimensionality.
- It removes irrelevant features from the data because having irrelevant features in the data can decrease accuracy of models and make your model learn based on irrelevant features.

AMEY B 50 Amey

Page No.:
Date:

$$A^T = \begin{bmatrix} 2 & 1 & 0 & -1 \\ 4 & 3 & 1 & 0.5 \end{bmatrix}$$

$$\Rightarrow \text{Covariance} = \sum_{i=1}^{n} \frac{(x-\bar{x})(y-\bar{y})}{n-1}$$

Lets consider upper row as x and lower row as y

Here $n = 4$.

x	y	$(x-\bar{x})$	$(y-\bar{y})$	$[(x-\bar{x})(y-\bar{y})]$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
2	4	1.5	1.875	2.8185	2.25	3.5156
1	3	0.5	0.875	0.4375	0.25	0.7656
0	1	-0.5	-1.125	0.5625	0.25	1.2656
-1	0.5	-1.5	-1.625	2.4375	2.25	2.6404

$$\bar{x} = \frac{2+1+0+(-1)}{4} = 0.5$$

$$\bar{y} = \frac{4+3+1+0.5}{4} = 2.125$$

Now,

$$\Sigma(x-\bar{x})\cdot(y-\bar{y}) = 2.8185 + 0.4375 + 0.5625 + 2.4375 = 6.256$$

$$\Sigma(x-\bar{x})^2 = 2.25 + 0.25 + 0.25 + 2.25 = 5$$

$$\Sigma(y-\bar{y})^2 = 3.5156 + 0.7656 + 1.2656 + 2.6404 = 8.1872$$

AMEY B 50 Amey

Page No.:

Date:

$$\text{Covariance } (x, x) = \sum_{i=1}^{n} \frac{(x-\bar{x})^2}{n-1} = \frac{5}{3} = 1.67$$

$$\text{Covariance } (y, y) = \sum_{i=1}^{n} \frac{(y-\bar{y})^2}{n-1} = \frac{8.1872}{3} = 2.73$$

$$\text{Covariance } (x, y) = \sum_{i=1}^{n} \frac{(x-\bar{x}).(y-\bar{y})}{n-1} = \frac{6.256}{3} = 2.09$$

Putting these values in a matrix form as S

$$S = \text{Covariance} = \begin{bmatrix} cov(x,x) & cov(x,y) \\ cov(y,x) & cov(y,y) \end{bmatrix}$$

$$= \begin{bmatrix} 1.67 & 2.09 \\ 2.09 & 2.73 \end{bmatrix}$$

Putting these values in characteristics equation

$$\begin{bmatrix} 1.67 & 2.09 \\ 2.09 & 2.73 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = 0$$

$$= \begin{vmatrix} 1.67 - \lambda & 2.09 \\ 2.09 & 2.73 - \lambda \end{vmatrix} = 0 \qquad - (1)$$

$$(1.67 - \lambda) \cdot (2.73 - \lambda) - (2.09)^2 = 0$$

$$\lambda^2 - 4.4\lambda + 0.191 = 0$$

$$\lambda_1 = 4.3562$$

$$\lambda_2 = 0.0439$$

(A) Taking λ_1 and putting in eq^n (1), to get a_{11} & a_{12}.

$$\begin{bmatrix} 1.67-4.3562 & 2.09 \\ 2.09 & 2.73-4.3562 \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{12} \end{bmatrix} = 0$$

$$\begin{bmatrix} -2.6862 & 2.09 \\ 2.09 & -1.6262 \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{12} \end{bmatrix} = 0$$

$$-2.6862 a_{11} + 2.09 a_{12} = 0$$
$$2.09 a_{11} - 1.6262 a_{12} = 0$$

$$\therefore a_{12} = \frac{2.6862}{2.09} a_{11}$$

Using equation of orthogonal transformation relation

$$a_{11}^2 + a_{12}^2 = 1$$

substituting values

$$a_{11}^2 + (1.2853 a_{11})^2 = 1$$

$$2.5620 a_{11}^2 = 1$$

$$a_{11} = 0.64$$

$$\therefore a_{12} = 0.7684 \quad [\because a_{11}^2 + a_{12}^2 = 1]$$

AMEY B 50 Amey

(B) Taking λ_2 in eq^n (1) to get a_{21} and a_{22}

$$\begin{bmatrix} 1.67-0.0439 & 2.09 \\ 2.09 & 2.73-0.0439 \end{bmatrix} \begin{bmatrix} a_{21} \\ a_{22} \end{bmatrix} = 0$$

$$\begin{bmatrix} 1.6261 & 2.09 \\ 2.09 & 2.6861 \end{bmatrix} \begin{bmatrix} a_{21} \\ a_{22} \end{bmatrix} = 0$$

$$1.6261 a_{21} + 2.09 a_{22} = 0$$
$$2.09 a_{21} + 2.6861 a_{22} = 0$$

$$\therefore a_{22} = \frac{1.6261}{2.09} a_{21}$$

$$= -0.7781 a_{21}$$

$$a_{21}^2 + a_{22}^2 = 1$$

Substituting values

$$a_{21}^2 + (-0.7781 a_{21})^2 = 1$$

$$a_{21} = 0.7893$$

$$\therefore a_{22} = 0.6141 \quad [\because a_{21}^2 + a_{22}^2 = 1]$$

AMEY B 50 Amey

Page No.: Date:

Therefore using values of a_{11}, a_{12}, a_{21}, a_{22}.

The principal components are :

$$Z_1 = a_{11} x_1 + a_{12} x_2$$

$$= 0.64 x_1 + 0.7684 x_2$$

$$Z_2 = a_{21} x_1 + a_{22} x_2$$

$$= 0.7893 x_1 + 0.6141 x_2$$

AMEY B 50 Amey

Page No.:
Date:

Q6. Describe the down Hill simplex method. Why it is Called the Derivative - Free method ?

Ans:

Downhill Simplex Method

- The downhill simplex method is an optimization algorithm.
- It doesn't make any assumption on the cost function to minimize.
- It is a conventional direct search algorithm where the best solution lies on the vertices of geometric figure in N dimensional space made of a set of N+1 points.
- The method compares the objective function values at the N+1 vertices and moves towards the optimum point.
- The movement of simplex algorithm is achieved by reflection, connection and expansion. It is called derivative - free optimization because it is used when it is difficult to find function derivatives or if finding such derivative is time consuming.

AMEY B 50 Amey

Page No.:
Date:

Q7. Minimize $f(x_1, x_2) = 4x_1 - 2x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$ With starting point $x_1 = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix}$ using the steepest decent method.

Ans:

Step 1:

We have

$$f(x_1, x_2) = 4x_1 - 2x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$$

The gradient of f is given by

$$\nabla f = \begin{Bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{Bmatrix} = \begin{Bmatrix} 4 + 4x_1 + 2x_2 \\ -2 + 2x_1 + 2x_2 \end{Bmatrix}$$

$$\nabla f = \nabla f(x_1) = \begin{Bmatrix} 4 \\ -2 \end{Bmatrix}$$

Step 2:

We have

$$S_i = -\nabla f_1 = -\nabla f(x_1)$$

$$\therefore S_i = -\nabla f_1 = \begin{Bmatrix} 4 \\ -2 \end{Bmatrix}$$

Step 3:

Determine λ_1^* in the direction S_i to find x_2. we need to find the optimal length λ_1 for this, we minimize.

$$f(x_1 + \lambda_1 S_1) \text{ with respect to } \lambda_1 :$$

AMEY B 50 Amey

Page No.:

Date:

$\therefore f(x_1 + \lambda_1 S_1) = f(-\lambda_1, \lambda_1) = \lambda_1^2 - 2\lambda_1$ — ①

$= x_1 + \lambda_1 S_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \lambda_1 \begin{bmatrix} -4 \\ 2 \end{bmatrix}$

As we have

$f(x_1, x_2) = 4x_1 + 2x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$

Hence

$f(-\lambda, \lambda_1) = -4\lambda_1 - 2\lambda_1 + 2\lambda_1^2 - 2\lambda_1\lambda_1 + \lambda_1^2$

$= -4\lambda_1 - 2\lambda_1 \cancel{+2\lambda_1^2} \cancel{-2\lambda_1^2} + \lambda_1^2$

$= \lambda_1^2 - 6\lambda_1$

Substituting derivatives with respect to λ_1

$\frac{\partial f}{\partial \lambda_1} = 0$

$\therefore \frac{\partial(\lambda_1^2 - 6\lambda_1)}{\partial \lambda_1} = 2\lambda_1 - 6 = 0$

$\therefore \lambda_1 = 3$

Now,

$X_2 = x_1 + \lambda_1 S_1 = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix} + 3 \begin{Bmatrix} -4 \\ 2 \end{Bmatrix} = \begin{Bmatrix} -12 \\ 6 \end{Bmatrix}$

As, $\nabla f_2 = \nabla f(x_2) \neq \begin{Bmatrix} 0 \\ 0 \end{Bmatrix}$

$\therefore X_2$ is not an optimal solution

To find optimal solution we have to perform several iterations

AMEY B 50 Amey

Page No.:
Date:

Q8. Find the singular value decomposition of $A = \begin{bmatrix} 2 & 2 \\ 1 & -1 \end{bmatrix}$

And also list some advantages of derivative based optimized techniques.

Ans:

$$A = \begin{bmatrix} 2 & 2 \\ 1 & -1 \end{bmatrix}$$

$$A^T = \begin{bmatrix} 2 & 1 \\ 2 & -1 \end{bmatrix}$$

Step 1

$$A^T . A = \begin{bmatrix} 5 & 3 \\ 3 & 5 \end{bmatrix}$$

Step 2

$$| A^T . A - \lambda |$$

$$= \begin{bmatrix} 5-\lambda & 3 \\ 3 & 5-\lambda \end{bmatrix}$$

$$= (5-\lambda)(5-\lambda) - 9 = 0$$

$$= \lambda^2 - 10\lambda + 16 = 0$$

$$\lambda = 8, 2$$

$$\therefore \lambda_1 = 8$$

$$\lambda_2 = 2$$

AMEY B 50 Amey

Page No.:

Date:

$S_1 = \sqrt{8} = 2.8284$

$S_2 = \sqrt{2} = 1.4142$

$\therefore S_1 > S_2$

Step 3:

$$S = \begin{bmatrix} 2.8284 & 0 \\ 0 & 1.4142 \end{bmatrix}$$

$$S^{-1} = \begin{bmatrix} 0.3535 & 0 \\ 0 & 0.7071 \end{bmatrix}$$

Step 4:

$\lambda_1 = 8$

$(A^T \cdot A - \lambda)(x_1) = 0$

$$\begin{bmatrix} 5-8 & 3 \\ 3 & 5-8 \end{bmatrix} x_1 = 0$$

$$\begin{bmatrix} -3 & 3 \\ 3 & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$$

$-3x_1 - 3x_2 = 0$

$3x_1 - 3x_2 = 0$

$x_2 = x_1$

$\lambda_2 = 2$

$(A^T \cdot A - \lambda) X_2 = 0$

$$\begin{bmatrix} 5-2 & 3 \\ 3 & 5-2 \end{bmatrix} x_2 = 0$$

$$\begin{bmatrix} 3 & 3 \\ 3 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$$

$3x_1 + 3x_2 = 0$

$3x_1 + 3x_2 = 0$

$x_2 = (-x_1)$

AMEY B 50 Amey

Page No.:
Date:

$\therefore X_1 = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_1 \end{bmatrix}$

$x_2 = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ -x_1 \end{bmatrix}$

Dividing by its length

$L = \sqrt{x_1^2 + x_2^2} = \sqrt{2x_1^2}$

$= x_1 \sqrt{2}$

$x_1 = \begin{bmatrix} x_1/L \\ x_2/L \end{bmatrix}$

$x_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$

$x_1 = \begin{bmatrix} 0.7071 \\ 0.7071 \end{bmatrix}$

Dividing by its length

$L = \sqrt{x_1^2 + x_2^2} = \sqrt{2x_1^2}$

$= x_1 \sqrt{2}$

$x_2 = \begin{bmatrix} x_1/L \\ -x_1/L \end{bmatrix}$

$x_2 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$

$x_2 = \begin{bmatrix} 0.7071 \\ -0.7071 \end{bmatrix}$

$V = [x_1 \cdot x_2]$

$= \begin{bmatrix} 0.7071 & 0.7071 \\ 0.7071 & -0.7071 \end{bmatrix}$

$V^T = \begin{bmatrix} 0.7071 & 0.7071 \\ 0.7071 & -0.7071 \end{bmatrix}$

AMEY B 50 Amey

Page No.:

Date:

Step 5 :

$$U = A V S^{-1}$$

$$U = \begin{bmatrix} 2 & 2 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 0.7071 & 0.7071 \\ 0.7071 & -0.7071 \end{bmatrix} \begin{bmatrix} 0.3535 & 0 \\ 0 & 0.7071 \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 2 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 0.25 & 0.5 \\ 0.25 & -0.5 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$A = U S V^T$$

$$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2.8284 & 0 \\ 0 & 1.4142 \end{bmatrix} \begin{bmatrix} 0.7071 & 0.7071 \\ 0.7071 & -0.7071 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 2 \\ 1 & -1 \end{bmatrix}$$

$$A = \begin{bmatrix} 2 & 2 \\ 1 & -1 \end{bmatrix}$$