MACHINE LEARNING ALGORITHMS

-S.DEEPAN KUMAR

TABLE OF CONTENTS

SUPPORT VECTOR MACHINE ALGORITHIM (SUPERVISED CLASSIFICATION)

- **GOAL – TO CREATE THE BEST LINE OR DECISION BOUNDARY THAT CAN SEGREGATE N-DIMENSIONAL SPACE INTO CLASSES SO THAT WE CAN EASILY PUT THE NEW DATA POINT IN THE CATEGORY IN THE FUTURE**

- **THE BEST DECISION BOUNDARY IS CALLED A HYPERPLANE**

- **SVM CHOOSES THE EXTREME POINTS/SUPPORT VECTORS THAT HELP IN CREATING THE HYPERPLANE**

- **THE DIMENSION OF THE HYPERPLANE DEPEND ON THE FEATURES PRESENT IN THE DATA SET**

- **OPTIMAL HYPERPLANE HAS THE MAXIMUM MARGIN (DISTANCE FROM THE SUPPORT VECTORS)**

- **RESULTS OF THE ALGORITHM ARE THE CLASSIFICATION OF A PARTICULAR DATA**

Support Vector Machine (SVM) Algorithm - Javatpoint

TYPES OF SUPPORT VECTOR MACHINE ALGORITHM

- 1.LINEAR SVM -

- ONLY WORKS ON LINEARLY SEPARABLE DATA

- (. I.E. DATA SET CAN BE CLASSIFIED BY A STRAIGHT LINE)

- 2.NONLINEAR SVM-

- WORKS ON NON-LINEARLY SEPARABLE DATA

- (.I.E. CANNOT BE CLASSIFIED BY USING A STRAIGHT LINE)
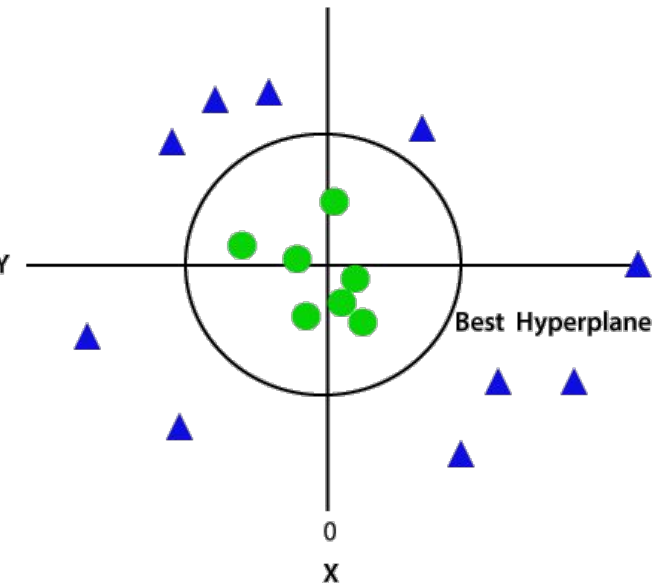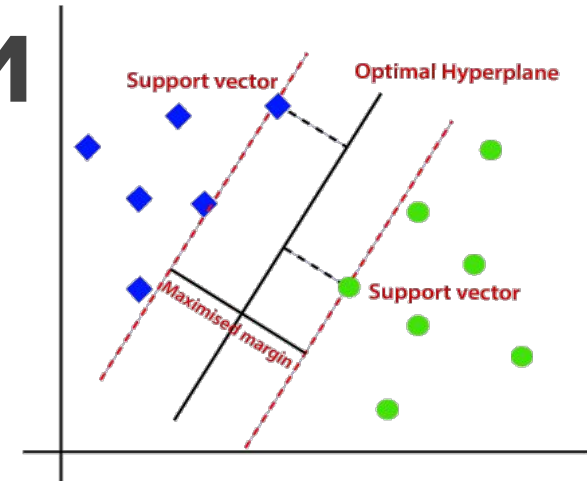
USES OF THIS ALGORITHM

1. Face detection
2. image classification
3. text categorization

FUZZY C MEANS CLUSTERING (UNSUPERVISED CLASSIFICATION)

- K-MEANS CLUSTERING (HARD CLUSTERING)

- FUZZY CLUSTERING(SOFT CLUSTERING)

1. INITIALIZE MEMBERSHIP VALUES

2. UPDATE CLASS CENTERS

3. CALCULATE DISTANCE OF SAMPLES FROM CLASS CENTERS

4. UPDATE MEMBERSHIP

WE USE EUCLIDIAN DISTANCE IN FUZZY C MEANS CLUSTERING

RANDOM FOREST

- THE RANDOM FOREST ALGORITHM IS MADE UP OF A COLLECTION OF DECISION TREES, AND EACH TREE IN THE ENSEMBLE IS COMPRISED OF A DATA SAMPLE DRAWN FROM A TRAINING SET WITH REPLACEMENT, CALLED THE BOOTSTRAP SAMPLE.

- COMBINING RESULTS WITH FORM MULTIPLE MODELS AGGREGATION

What is Random Forest? | IBM

APPLICATIONS OF RANDOM FOREST

- THE RANDOM FOREST ALGORITHM HAS BEEN APPLIED ACROSS A NUMBER OF INDUSTRIES, ALLOWING THEM TO MAKE BETTER BUSINESS DECISIONS. SOME USE CASES INCLUDE:

1. FINANCE: IT IS A PREFERRED ALGORITHM OVER OTHERS AS IT REDUCES TIME SPENT ON DATA MANAGEMENT AND PRE-PROCESSING TASKS. IT CAN BE USED TO EVALUATE CUSTOMERS WITH HIGH CREDIT RISK, TO DETECT FRAUD, AND OPTION PRICING PROBLEMS.

2. HEALTHCARE: THE RANDOM FOREST ALGORITHM HAS APPLICATIONS WITHIN COMPUTATIONAL BIOLOGY, ALLOWING DOCTORS TO TACKLE PROBLEMS SUCH AS GENE EXPRESSION CLASSIFICATION, BIOMARKER DISCOVERY, AND SEQUENCE ANNOTATION. AS A RESULT, DOCTORS CAN MAKE ESTIMATES AROUND DRUG RESPONSES TO SPECIFIC MEDICATIONS.

3. E-COMMERCE: IT CAN BE USED FOR RECOMMENDATION ENGINES FOR CROSS-SELL PURPOSES.

RECURRENT NEURAL NETWORKS

Rolled RNN

Unrolled RNN

Output layer

Y_t $=$ Y_0 Y_1 Y_2

Hidden layers

Input layer

X_t X_0 X_1 X_2
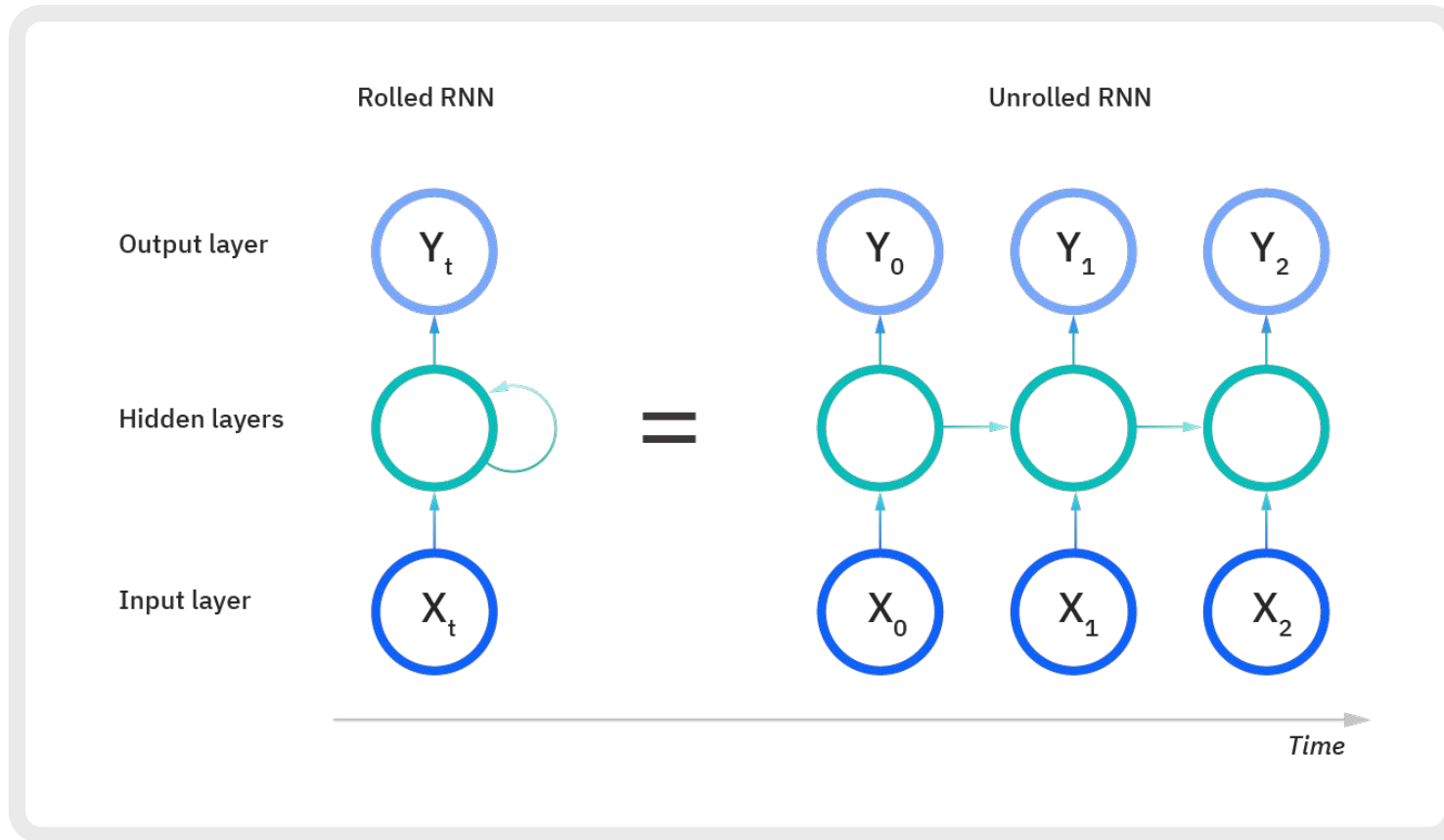
Time

- A RECURRENT NEURAL NETWORK (RNN) IS A TYPE OF ARTIFICIAL NEURAL NETWORK WHICH USES SEQUENTIAL DATA OR TIME SERIES DATA. THESE DEEP LEARNING ALGORITHMS ARE COMMONLY USED FOR ORDINAL OR TEMPORAL PROBLEMS, SUCH AS LANGUAGE TRANSLATION, NATURAL LANGUAGE PROCESSING (NLP), SPEECH RECOGNITION, AND IMAGE CAPTIONING

- THEY ARE DISTINGUISHED BY THEIR "MEMORY" AS THEY TAKE INFORMATION FROM PRIOR INPUTS TO INFLUENCE THE CURRENT INPUT AND OUTPUT.

What are Recurrent Neural Networks? | IBM

THANK YOU