1. What is NB Tree.

Ans: It's a hybrid learning approach of decision tree & naive bayes classifier.

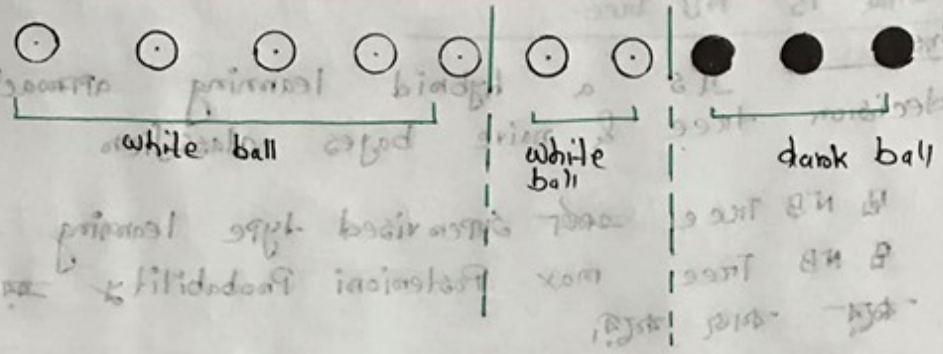
- NB Tree একটা supervised type learning
- NB Tree max Posterioni Probability এর উপর base করে কাজ করে।
- NB Tree = Decision Tree + NB
- Decision tree এর leaf node replaced এর NB model দিয়ে।
- NB Tree একটা Pruned tree.
- Pruned tree তৈরিতে NB tree majority class নিয়ে কাজ করে,
- leaf গুলো একটা sub-data set নিয়ে কাজ করে,

2. Why NB Tree is useful?

Ans:

Adaptive NB Tree splits the dataset by applying entropy based algorithm & then used standard NB classifier at the leaf node to handle attribute.

It applies strategy to construct decision tree & replaces leaf node with NB classifier.



Joint Probability

For (White-ball) = $\frac{7}{10}$

For (dark-ball) = $\frac{3}{10}$

5 white balls together = $\frac{5}{5} = 1$ // Highest Probability

2 " " " = $\frac{2}{2}$

3 dark " = $\frac{3}{5}$

Main Motivation: We if divide dataPoint, dataset in such manner, the same would be together.

3. Define One-R Classifier

Ans: It's a classification rule. Based on the value of a single Predictor, that generates one rule for each Predictor in the data.

One-R classifier এর জন্য frequency table construct করতে হয়।

4. What is Frequency Table

Ans: It's a table that represents the number of occurences of every unique value in the variable.

5. Write the Algorithm of One R Classifier

Ans:

Input : $D = \{x_1, x_2, \ldots, x_n\}$ // Training Data
Output : OneR Model.
Method

```
for each attribute Ai ∈ D, do
    for each attribute value Aij ∈ Ai, do
        Make a classifier rule:
            Count how often each class appears
            find the most frequent class
            make the rule assign that class to this Aij
    end for
    Calculate the error rate of this attribute's Ai rule
end for
choose the attribute Ai ∈ D with the smallest error rate
```

Table: The playing tennis dataset

Day	Outlook	Temperature	Humidity	Wind	Play
D_1	Sunny	Hot	High	Weak	No
D_2	Sunny	Hot	High	Strong	No
D_3	Overcast	Hot	High	Weak	Yes
D_4	Rain	Mild	High	Weak	Yes
D_5	Rain	Cool	Normal	Weak	Yes
D_6	Rain	Cool	Normal	Strong	No
D_7	Overcast	Cool	Normal	Strong	Yes
D_8	Sunny	Mild	High	Weak	No
D_9	Sunny	Cool	Normal	Weak	Yes
D_{10}	Rain	Mild	Normal	Weak	Yes
D_{11}	Sunny	Mild	Normal	Strong	Yes
D_{12}	Overcast	Mild	High	Strong	Yes
D_{13}	Overcast	Hot	Normal	Weak	Yes
D_{14}	Rain	Mild	High	Strong	No

- Inner loop দ্বারা Classification rule বানানো.
- Outer loop দ্বারা প্রতিটা attribute এর জন্য error rate হিসেব করছে.

Simulation of One R Based on Playing Tennis Dataset

Attribute ⇒ Outlook

Item	Attribute Value		Error	Total Error
1	Sunny	Yes = 2/5 = 0.4	2/5	$\frac{2}{5}+\frac{0}{4}+\frac{2}{5}$
		no = 3/5 = 0.6		
2	Overcast	Yes = 4/4 = 1	0/4	$=\frac{2+0+2}{5+4+5}$
		no = 0/4 = 0		$=\frac{4}{14}$
3	Rain	Yes = 3/5 = 0.6	2/5	
		no = 2/5 = 0.4		

* মোট error
* $\frac{4}{14}$ মানে total 4 টা miss classified

Attribute => Temperature

Item	Attribute Value		Entropy	Total Entropy
1	Hot	yes = 2/4 = 0.5	2/4	$\frac{2}{4}+\frac{3}{6}+\frac{1}{4}$
		no = 2/4 = 0.5		$=\frac{2+2+1}{4+6+4}$
2	Mild	yes = 4/6 = 0.6	2/6	$=\frac{5}{14}$
		no = 2/6 = 0.3		
3	Cool	yes = 3/4 = 0.75	1/4	
		no = 1/4 = 0.25		

যেখানে কম সেখানে একটা দিব।
Less Entropy যেটা সেটা ভালো।

Attribute => Humidity

Item	Attribute Value		Entropy	Total Entropy
1	High	yes = 3/7 = 0.4	3/7	$\frac{3}{7}+\frac{1}{7}$
		no = 4/7 = 0.5		$=\frac{3+1}{7+7}$
2	Normal	yes = 6/7 = 0.8	1/7	$=\frac{4}{14}$
		no = 1/7 = 0.1		

Attribute => Wind

Item	Attribute Value		Entropy	Total Entropy
1	Weak	yes = 6/8 = 0.7	2/8	$\frac{2}{8}+\frac{3}{6}$
		no = 2/8 = 0.2		$=\frac{2+3}{8+6}$
2	Strong	yes = 3/6 = 0.5	3/6	$=5/14$
		no = 3/6 = 0.5		

Entropy Rate কম থাকলে accuracy বেশি

6. What is Support Vector Machine

Ans:

It's a supervised learning model, with associated learning algorithms that analyze data used for classification & regression analysis.

SVM training algorithms build a model that assigns new vectors into one category or the other.

7. Write the working Process of SVM

Ans:

It's a representation of the vectors as point in space, mapped so that the vectors of the separate categories are divided by a clear gap that is as wide as possible.

New vectors are then mapped into that same space & predicted to belong to a category based on which side of the gap.

SVM non-linear classification perform করে থাকে, এজন্যে এরা হয় kernel trick.

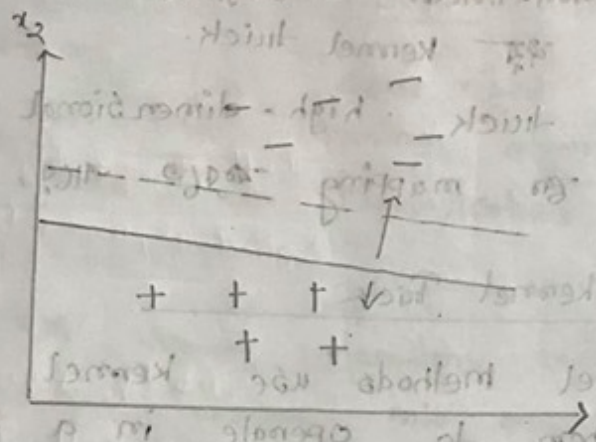
kernel trick high-dimensional feature space input কে mapping করতে পারে,

8. What is kernel Trick

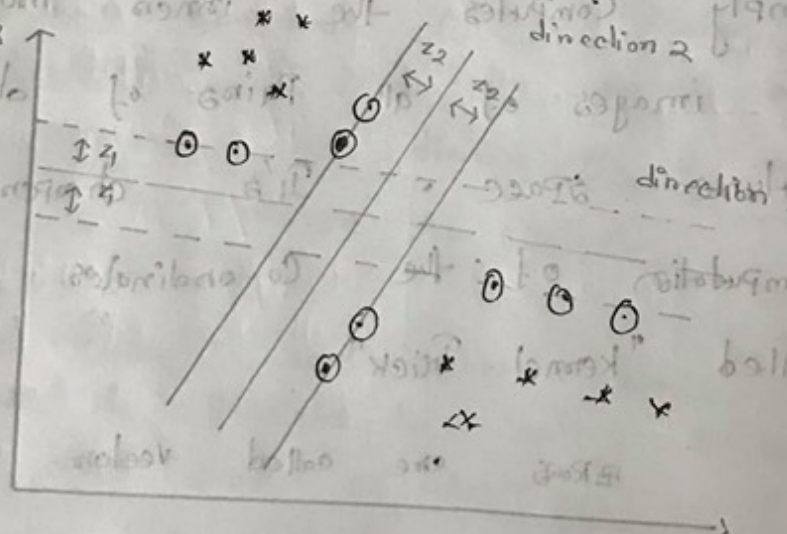
Ans:

Kernel methods use kernel functions. It enable them to operate in a high-dimensional implicit feature space without even computing the co-ordinates of the data in that space. It simply computes the inner product between the images of all pairs of data in the feature space. It's cheaper than explicit computation of the co-ordinates. This approach is called "kernel Trick".

Rows are called vectors.



-> এটা একটা linearly separable two class Problem.



-> এটা একটা linearly separable two class Problem with two possible linear classifiers

-> ⊙ এটা দ্বারা instance গুলো miss classified

9. Show the relation between vector inner Product & norm with Projection.

Ans.

Let,
we have two dimensional vectors

$$u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \quad \& \quad v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

$$u^T v = \begin{bmatrix} u_1 & u_2 \end{bmatrix} \begin{bmatrix} v_2 \\ v_1 \end{bmatrix} = v^T u = P \cdot \|u\|$$

$$\|u\| = \sqrt{u_1^2 + u_2^2} \qquad [u_1, u_2 \in \mathbb{R}]$$

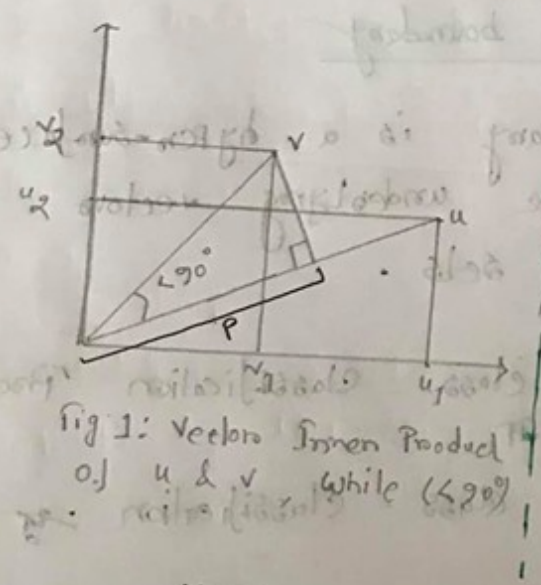


Fig 1: Vector Inner Product of u & v while (<90°)

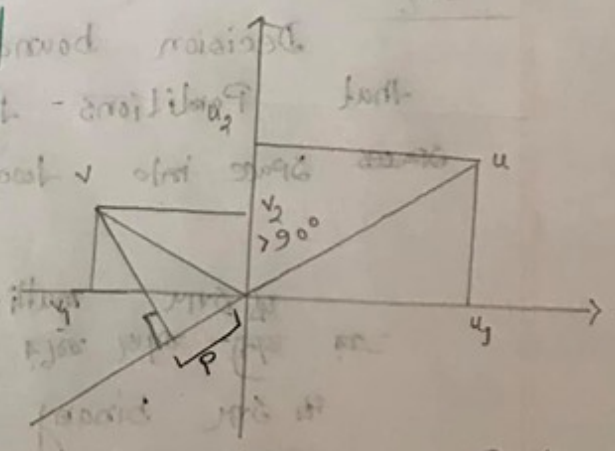


Fig: Vector Inner Product of u & v, while (>90°)

Here

The angles find support vector.
P means vector inner Product
u_1 = x value of u
u_2 = y value of u
v_1 = x value of v
v_2 = y value of v

The greater the ~~gap~~ P, ~~the~~ the gap is maximized

So,

$$u^T v = P \cdot \|u\|$$

$$u^T v = u_1 v_1 + u_2 v_2$$

So,

$$u^T v = v^T \cdot u.$$

$$u^T v = P \cdot \|u\| = v^T u = u_1 v_1 + u_2 v_1 \leftarrow P \in \mathbb{R}.$$

Minimizing θ, maximizes P.

10. Define SVM decision boundary

Ans.

Decision boundary is a hyper-surface that Partitions the underlying vector ~~spaces~~ space into two sets.

SVM multi-class classification Problem এর জন্য কাজে লাগে না,

SVM binary class - classification এর জন্য কাজে লাগে।

11. How to overcome Slow Processing of SVM

Ans:

Kernel trick is the solution. In kernel trick mapping the non-linear separable data-set into a higher dimensional space where we can find a hyperplane that can separate the sample.