Name	Gender	Height	Output
Kristina	F	1.6 m	Short
Jim	M	2m	Tall
Maggie	F	1.9 m	Medium
Martha	F	1.88 m	Medium
Stephanie	F	1.7m	Short
Bob	M	1.85 m	Medium
Kathy	F	1.6 m	Short
Dave	M	1.7 m	Short
Worth	M	2.2 m	Tall
Steven	M	2.1m	Tall
Debbie	F	1.8 m	Medium
Todd	M	1.95 m	Medium
kim	F	1.9 m	Medium
Amy	F	1.8 m	Medium
Wynette	F	1.75 m	Medium

(Pat, F, 1.6) Classify it, while k=5

Name	Height	Distance	Nearest N	Majority class	Majority Voting
Kristina	1.6m	$\sqrt{(1.6-1.6)^2} = 0$	Yes	Short	
Jim	2m	$\sqrt{(2-1.6)^2} = 0.4$	No		
Maggie	1.9m	$\sqrt{(1.9-1.6)^2} = 0.3$	No		
Martha	1.88m	$\sqrt{(1.88-1.6)^2} = 0.28$	No		
Stephanie	1.7m	$\sqrt{(1.7-1.6)^2} = 0.1$	Yes	Short	
Bob	1.85m	$\sqrt{(1.85-1.6)^2} = 0.25$	~~Yes~~ No		
Kathy	1.6m	$\sqrt{(1.6-1.6)^2} = 0$	Yes	Short	Short.
Dave	1.7m	$\sqrt{(1.7-1.6)^2} = 0.1$	Yes	~~Medium~~ Short	
Worth	2.2m	$\sqrt{(2.2-1.6)^2} = 0.6$	No		
Steven	2.1m	$\sqrt{(2.1-1.6)^2} = 0.5$	No		
Debbie	1.8m	$\sqrt{(1.8-1.6)^2} = 0.2$	Yes	~~Tall~~ Medium	
Todd	1.95m	$\sqrt{(1.95-1.6)^2} = 0.35$	No		
Kim	1.9m	$\sqrt{(1.9-1.6)^2} = 0.3$	No		
Amy	1.8m	$\sqrt{(1.8-1.6)^2} = 0.2$	No		
Wynette	1.75m	$\sqrt{(1.75-1.6)^2} = 0.15$	No		

So,

KNN will classify Pat as Short.

- KNN ⇒ similarities নিয়ে কাজ করে
- Naive Bayes Classifier ⇒ Probability নিয়ে কাজ করে
- Naive means "simple"

Classifying Medical Image

- Grey Color image হলে Non malignant tumor [ক্যান্সার নয়]
- Dark color image হলে Malignant [ক্যান্সার]
- Image এর Pixel গুলোকে 0,1 দিয়ে represent করা যায়, যেমন

→ Pixel dark মানে : 1
→ Pixel grey মানে : 0

অথবা

→ Pixel grey মানে : 0
→ Pixel black মানে : 1

- Draw block diagram of classification model

Ans:

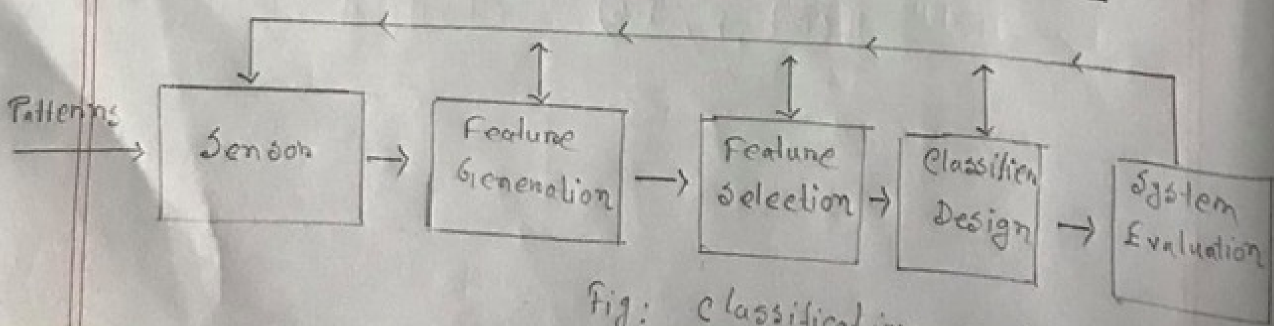


Fig: Classification Model

□ System Evaluation এ
→ Accuracy check
এবং
→ Performance Test হয়।

□ Define Feature Engineering / Feature Generation

Ans:

The Process of using domain knowledge of the data to create features that make machine learning algorithms work.

□ Image classification এর ক্ষেত্রে image থেকে feature extract করা হয়, সোজা কথায় image কে
□ matrix এ convert করা হয়,

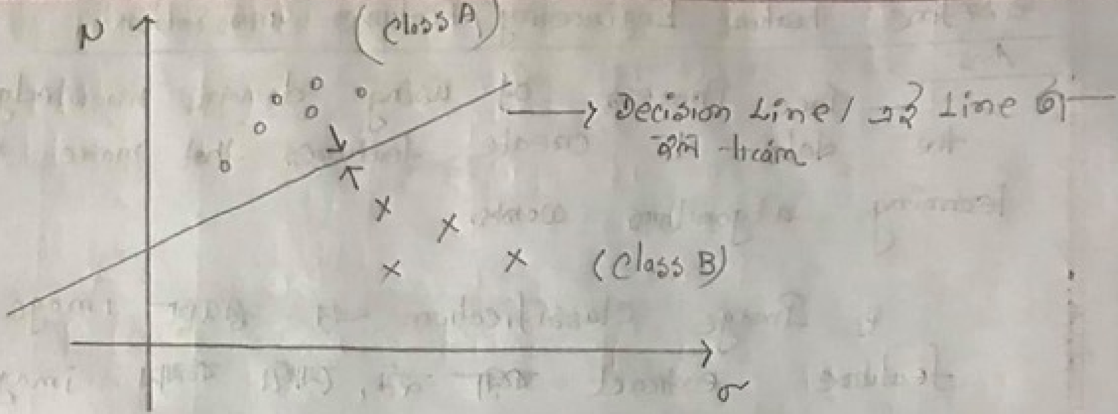
□ Features হতে পারে
→ Mean (μ)
→ Standard Deviation (σ)
→ Number of Zeros
→ Numbers of Ones
→ Average of 0's & Ones.

□ Write differences between KNN vs Naive Bayes Classifier

Ans:

KNN	Naive Bayes Classifier
a. High variance & low bias	a. low Variance & high bias.
b. Based on similarities	b. Based on Probability
c. No decision line.	c. It has decision line

□ KNN কে lazy learner বলার অন্যতম কারন হল এটা decision line আঁকতে পারে না।
□ Feature Generation করা মানে Parameter ঠিক করা
□ Parameter মানে column.



Decision line থেকে class গুলোর gape যত বেশি, তত ভালো, Gape বেশি মানে class গুলো biased না,

Define decision line

Ans:

Decision line divide the feature space into regions that corresponds to different classes.

Priori Probability Historical Data এর উপর কাজ করে,

Naive Bayes Classifier Priori Probability নিয়ে কোন decision দিতে পারে না, এর জন্য Class Conditional Probability দরকার হয়,

- M হলো ~~Class~~ Features এর সংখ্যা, M=1 হলে Classification হবে না।

- ধরি,

  $M = \{\omega_1, \omega_2\}$

  $\omega_1 \rightarrow$ Non Cancer

  $\omega_2 \rightarrow$ Cancer

  $N \Rightarrow$ Total Training data.

- $P(\omega_i \mid X)$

  $\omega_i \rightarrow$ Class

  $X \rightarrow$ Test data.

  Class values of X

- Naive Bayes এর ফর্মুলা হল

$$P(\omega_i \mid X) = \frac{P(X \mid \omega_i)\, P(\omega_i)}{P(X)}$$

$P(X \mid \omega_i) \rightarrow$ class conditional Probability

$P(\omega_i) \rightarrow$ Priori Probability

→ নিচের P(X) কে omit করা হয়, কারণ P(X) হলো সব class এর জন্য Constant.

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

Prior Probability:

$P(Play = Yes) = \frac{9}{14} = 0.642$

$P(Play = No) = \frac{5}{14} = 0.375$

Parameter → Outlook

$P(outlook = Sunny \mid Play = Yes) = \frac{2}{P(Play = Yes)}$

$= \frac{2}{9}$

$= 0.222$

$P(outlook = Sunny \mid Play = No) = \frac{3}{P(Play = No)}$

$= \frac{3}{5}$

$= 0.6$

$P(outlook = Overcast \mid Play = Yes) = \frac{4}{P(Play = Yes)}$

$= \frac{4}{9}$

$= 0.444$

$P(outlook = Overcast \mid Play = No) = \frac{0}{P(Play = No)}$

$= \frac{0}{5}$

$= 0$

$P(outlook = Rain \mid Play = Yes) = \frac{3}{P(Play = Yes)}$

$= \frac{3}{9}$

$= 0.3$

$P(outlook = Rain \mid Play = No) = \frac{2}{P(Play = No)} = \frac{2}{5} = 0.4$

Parameter⇒Temparature

$P(\text{Temparature} = \text{Hot} \mid \text{Play} = \text{Yes}) = \frac{2}{P(\text{Play}=\text{Yes})} = \frac{2}{9} = 0.222$

$P(\text{Temparature} = \text{Hot} \mid \text{Play} = \text{No}) = \frac{2}{P(\text{Play}=\text{No})} = \frac{2}{5} = 0.4$

$P(\text{Temparature} = \text{Mild} \mid \text{Play} = \text{Yes}) = \frac{4}{P(\text{Play}=\text{Yes})} = \frac{4}{9} = 0.444$

$P(\text{Temparature} = \text{Mild} \mid \text{Play} = \text{No}) = \frac{2}{P(\text{Play}=\text{No})} = \frac{2}{5} = 0.4$

$P(\text{Temparature} = \text{Cool} \mid \text{Play} = \text{Yes}) = \frac{3}{P(\text{Play}=\text{Yes})} = \frac{3}{9} = 0.333$

$P(\text{Temparature} = \text{Cool} \mid \text{Play} = \text{No}) = \frac{1}{P(\text{Play}=\text{No})} = \frac{1}{5} = 0.2$

Parameter ⇒ Humidity

$P(\text{Humidity} = \text{High} \mid \text{Play} = \text{Yes}) = \frac{3}{P(\text{Play}=\text{Yes})} = \frac{3}{9} = 0.333$

$P(\text{Humidity} = \text{High} \mid \text{Play} = \text{No}) = \frac{4}{P(\text{Play}=\text{No})} = \frac{4}{5} = 0.8$

$P(\text{Humidity} = \text{Normal} \mid \text{Play} = \text{Yes}) = \frac{6}{P(\text{Play}=\text{Yes})} = \frac{6}{9} = 0.666$

$P(\text{Humidity} = \text{Normal} \mid \text{Play} = \text{No}) = \frac{1}{P(\text{Play}=\text{No})} = \frac{1}{5} = 0.2$

Parameter ⇒ Wind

$P(\text{Wind} = \text{Weak} \mid \text{Play} = \text{Yes}) = \frac{6}{P(\text{Play}=\text{Yes})} = \frac{6}{9} = 0.666$

$P(\text{Wind} = \text{Weak} \mid \text{Play} = \text{No}) = \frac{2}{P(\text{Play}=\text{No})} = \frac{2}{5} = 0.4$

$P(\text{Wind} = \text{Strong} \mid \text{Play} = \text{Yes}) = \frac{3}{P(\text{Play}=\text{Yes})} = \frac{3}{9} = 0.333$

$P(\text{Wind} = \text{Strong} \mid \text{Play} = \text{No}) = \frac{3}{P(\text{Play}=\text{No})} = \frac{3}{5} = 0.6$

$P(D_1 \mid Play = Yes) = P(Outlook = Sunny \mid Play = Yes) \times$
$P(Temparature = Hot \mid Play = Yes) \times$
$P(Humidity = High \mid Play = Yes) \times P(Wind = Weak \mid Play = Yes)$

$= 0.222 \times 0.222 \times 0.333 \times 0.666$

$= 0.0109$

$P(D_1 \mid Play = No) = P(Outlook = Sunny \mid Play = No) \times$
$P(Temparature = Hot \mid Play = No) \times$
$P(Humidity = High \mid Play = No) \times P(Wind = Weak \mid Play - No)$

$= 0.6 \times 0.4 \times 0.8 \times 0.4$

$= 0.0768$

Now Classifying

$P(D_1 \mid Play = Yes) \times P(Play = Yes) = 0.0109 \times 0.642$
$= 0.00699$

$P(D_1 \mid Play = No) \times P(Play = No) = 0.0768 \times 0.375$
$= 0.0288$

$\underset{(No)}{0.0288} > \underset{(Yes)}{0.00699}$

So, Instance D_1 would be classified (Predicted) Play = No.