Data Science and Artificial Intelligence

Machine Learning

Bayesian Learning

Lecture No. 5

By- SIDDHARTH SABHARWAL SIR

Recap of Previous Lecture

Topic
- Continuous nature dimension
  ↳ Probab $\underline{N(\mu, \sigma^2)}$

Topic
- laplace smoothing

Topic
- Probab of error

Topic

Topic

Topics to be Covered

- Topic
- Topic
- Topic
- Topic
- Topic

- Type of models.
- Questions
- Person
- No of Parameters in Naive Bayes.

ONE SMALL

POSITIVE

THOUGHT

IN THE

MORNING

CAN CHANGE YOUR

WHOLE DAY

Naïve Bayes Classifier

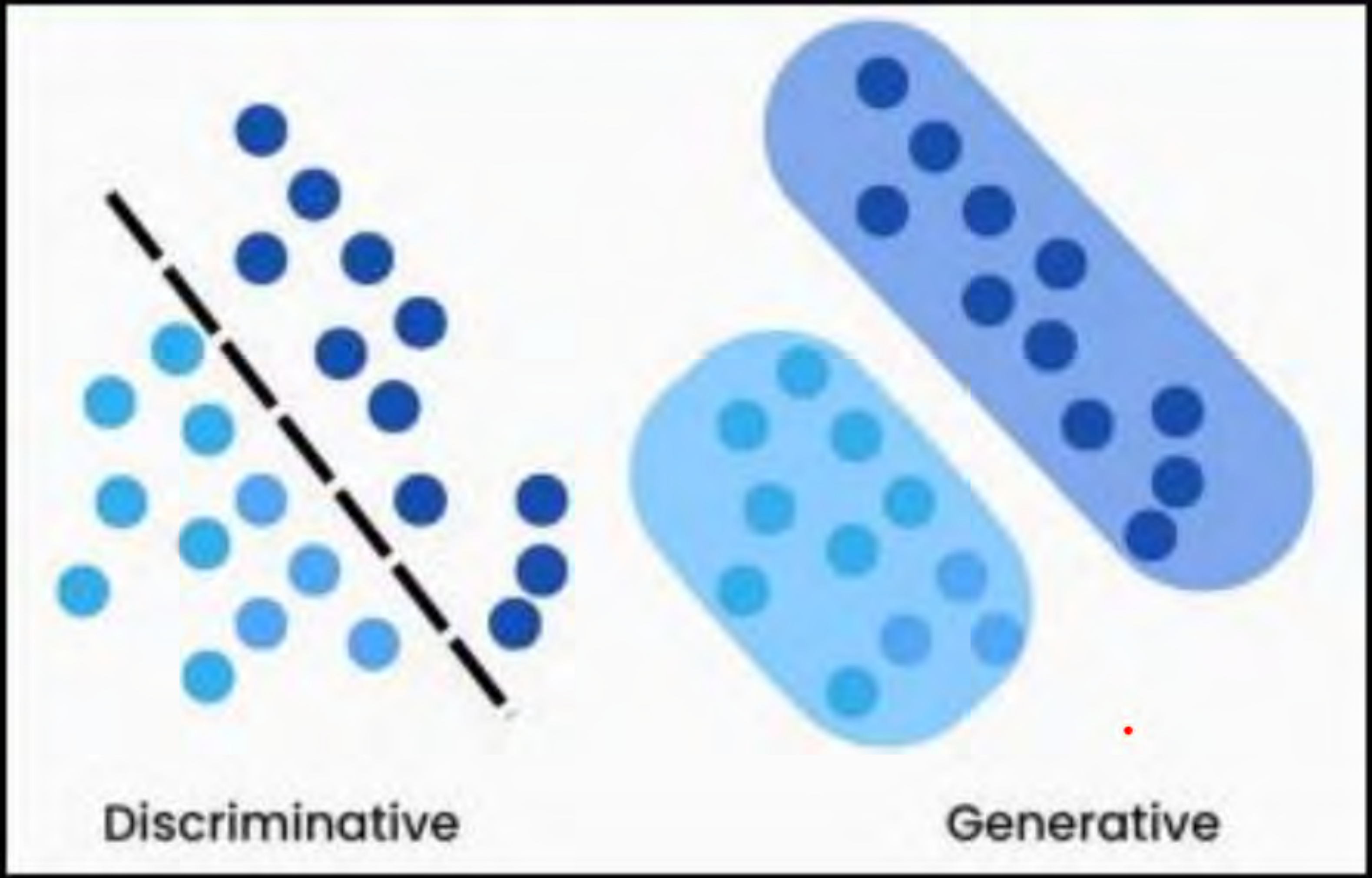
Advantages of Naïve Bayes Classifier:

- Naïve Bayes is one of the fast and easy ML algorithms to predict a class of datasets.
- It can be used for Binary as well as Multi-class Classifications.
- It performs well in Multi-class predictions as compared to the other Algorithms.
- It is the most popular choice for **text classification problems.**

Disadvantages of Naïve Bayes Classifier:

- Naive Bayes assumes that all features are independent or unrelated, so it cannot learn the relationship between features.
- Can be influenced by irrelevant attributes.
- May assign zero probability to unseen events, leading to poor generalization.

Discriminative vs. Generative Learning



Naive Bayes = Only for Classification

Papa

Question

Kid1
extract features

Clanifier use karke
find animal

Kid 1

Zoo

Kuch features
Height
weight
Colour
Tail length

Classifier
Banaliya

Kid2

more detail needed.
more data needed

Sabhi animals
Ki images
yaad Karli

and then compare
to predict
animal, whenever
Question asked

Bayesian Decision Theory

Story...

depth

A father has two kids, Kid A and Kid B. Kid A has a special character whereas he can learn everything in depth. Kid B have a special character whereas he can only learn the differences between what he saw.

One fine day, The father takes two of his kids (Kid A and Kid B) to a zoo. This zoo is a very small one and has only two kinds of animals say a lion and an elephant. After they came out of the zoo, the father showed them an animal and asked both of them "**is this animal a lion or an elephant**?"

The Kid A, the kid suddenly draw the image of lion and elephant in a piece of paper based on what he saw inside the zoo. He compared both the images with the animal standing before and answered based on the **closest match** of image & animal, he answered: "The animal is Lion".

The Kid B knows only the differences, based on **different properties learned**, he answered: "The animal is a Lion".

Here, we can see both of them is finding the kind of animal, but the way of learning and the way of finding answer is entirely different. In Machine Learning, We generally call Kid A as a Generative Model & Kid B as a Discriminative Model.

we find
data ka
Probab. distribution. ← $\boxed{\text{Naive Bayes}}$

we find
$P(x/\text{Clan}1)$ } → distribution of Clan 1
$P(x|\text{Clan}2)$ } → distribution of clan 2

Kid 2 ← Kid A ⇒ Generative model

Kid 1 ← Kid B ⇒ $\boxed{\text{discriminative model}}$ ← { Logistic Reg, Linear classification }

Discriminative vs. Generative Learning

Let's consider an example.
Imagine yourself as a language classification system.

There are two ways you can classify languages.

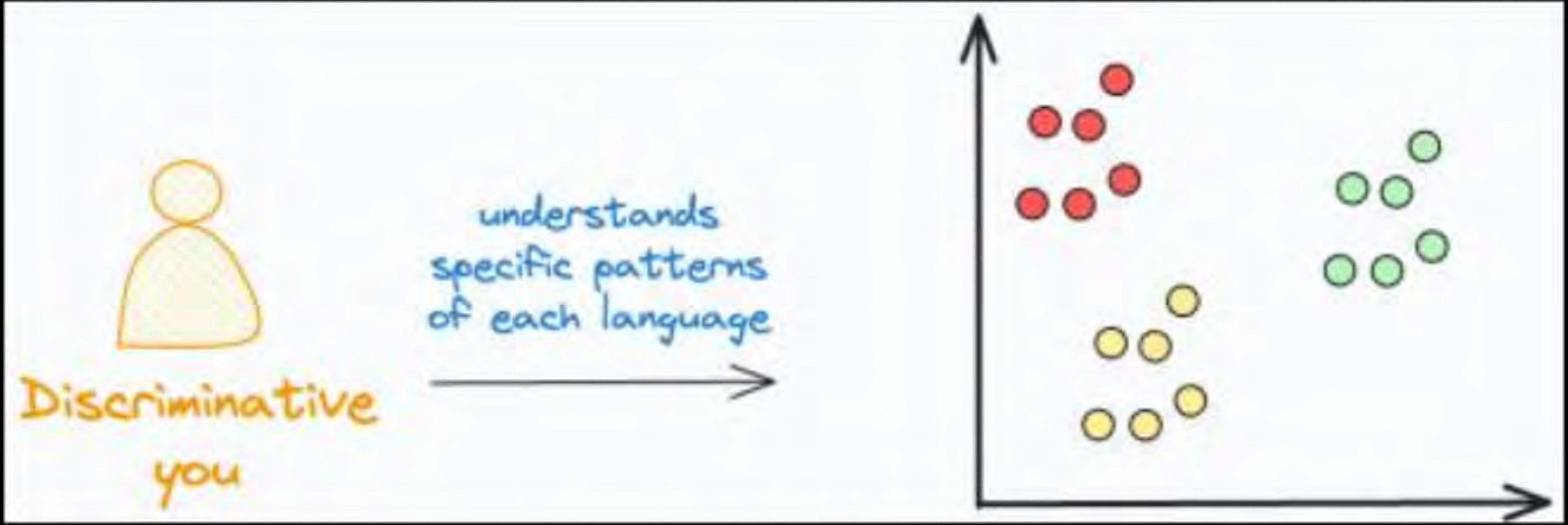
- Learn every language and then classify a new language based on acquired knowledge.
- Understand some distinctive patterns in each language without truly learning the language. Once done, classify a new language.

Can you figure out which of the above is generative and which one is discriminative?

Discriminative vs. Generative Learning

The second approach is a **discriminative approach**. This is because you only learned specific distinctive patterns of each language. It is like:

- If so and so words appear, it is likely "Langauge A."
- If this specific set of words appear, it is likely "Langauge B." and so on.



In other words, you learned the conditional distribution P(Language|Words).

Discriminative vs. Generative Learning

- Also, the above description might persuade you that generative models are more generally useful, but it is not true.
- This is because generative models have their own modeling complications.
- For instance, typically, generative models require more data than discriminative models.
- Relate it to the language classification example again.
- Imagine the amount of data you would need to learn all languages (generative approach) vs. the amount of data you would need to understand some distinctive patterns (discriminative approach).
- Typically, discriminative models outperform generative models in classification tasks.

Bayesian Decision Theory

LR|LogR

Naive Bayes.

Discriminative vs. Generative Learning

- In General, A Discriminative model models the **decision boundary between the classes.**
- A Generative Model explicitly models the **actual distribution of each class.**
- In final both of them is predicting the conditional probability P(Animal | Features). But Both models learn different probabilities.
- A Generative Model learns the **joint probability distribution p(x,y).** It predicts the conditional probability with the help of **Bayes Theorem.**
- A Discriminative model learns the **conditional probability distribution p(y|x).** Both of these models were generally used in **supervised learning** problems.

Discriminative Learning

- The discriminative model learn the boundaries between classes or labels in a dataset.
- Discriminative models focus on modelling the decision boundary between classes in a classification problem. The goal is to learn a function that maps inputs to binary outputs, indicating the class label of the input.
- Maximum likelihood estimation is often used to estimate the parameters of the discriminative model, such as the coefficients of a logistic regression model or the weights of a neural network.
- Discriminative models (just as in the literal meaning) separate classes. But these models are not capable of generating new data points. Therefore, the ultimate objective of discriminative models is to separate one class from another.
- If we have some outliers present in the dataset, discriminative models work better compared to generative models i.e., discriminative models are more robust to outliers.
- But overall the accuracy of discriminative model is less than the generative models.

Generative and Descriptive Learning

- Examples of Discriminative Models
  - Logistic regression
  - Support vector machines(SVMs)
  - Traditional neural networks
  - Nearest neighbor
  - Conditional Random Fields (CRFs)
  - Decision Trees and Random Forest
- Outliers have little to no effect on these models. They are a better choice than generative models, but this leads to misclassification problems which can be a major drawback.

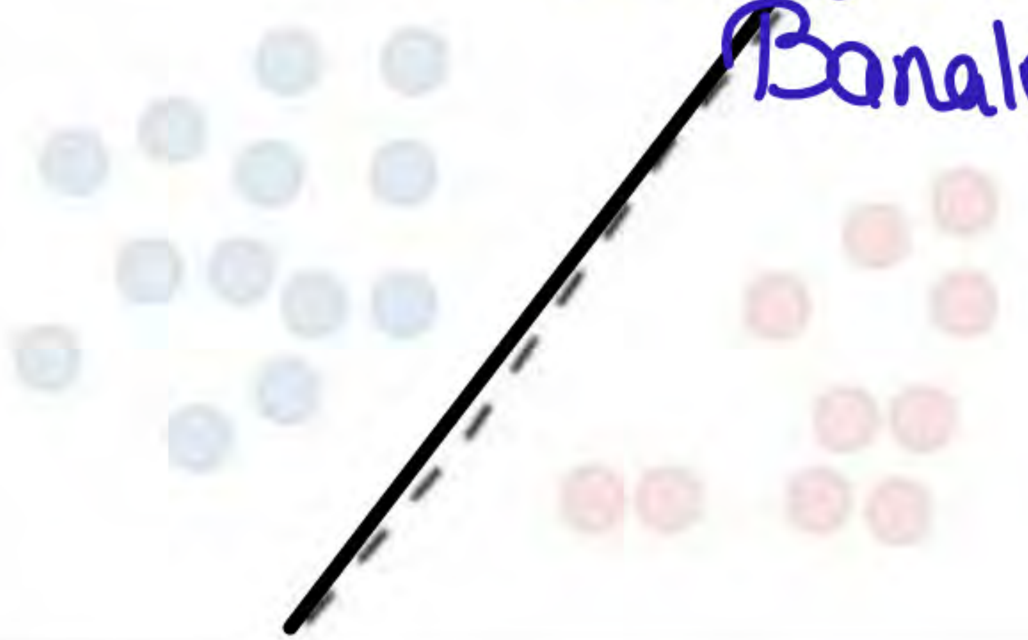
Generative Learning

- Generative models are machine learning models that learn to generate new data samples similar to the training data they were trained on. They capture the underlying distribution of the data and can produce novel instances.
- So, the Generative approach focuses on the distribution of individual classes in a dataset, and the learning algorithms tend to model the underlying patterns or distribution of the data points (e.g., gaussian). These models use the concept of joint probability and create instances where a given feature (x) or input and the desired output or label (y) exist simultaneously.
- These models use probability estimates and likelihood to model data points and differentiate between different class labels present in a dataset. Unlike discriminative models, these models can also generate new data points.
- However, they also have a major drawback – If there is a presence of outliers in the dataset, then it affects these types of models to a significant extent.

Generative and Descriptive Learning

- Generative model
- As the name suggests, generative models can be used to generate new data points. These models are usually used in unsupervised machine learning problems.
- Generative models go in-depth to model the actual data distribution and learn the different data points, rather than model just the decision boundary between classes.
- These models are prone to outliers, which is their only drawback when compared to discriminative models. The mathematics behind generative models is quite intuitive too. The method is not direct like in the case of discriminative models. To calculate $P(Y|X)$, they first estimate the prior probability $P(Y)$ and the likelihood probability $P(X|Y)$ from the data provided.

Generative and Descriptive Learning

	Discriminative model	Generative model
Goal	Directly estimate $P(y\|x)$	Estimate $P(x\|y)$ to then deduce $P(y\|x)$
What's learned	Decision boundary	Probability distributions of the data
Illustration	 Classifier Banalo	 distribution of Class 1/Class 0 data
Examples	Regressions, SVMs	GDA, Naive Bayes

* Number of Parameter in naive Bayes

D_1 Can take 4 values

we have to find $\boxed{P(D_1/C_1)}^{\to 4}$, $\boxed{P(D_2/C_1)}^{\to 3}$, $\boxed{P(D_3|C_1)}^{\to 4}$ ... $P(D_{10}/C_1)$

5 class $\boxed{P(D_1/C_2)}^{\to 4}$, - - - - - $P(D_{10}|C_2)$

$\boxed{P(D_1|C_5)}^{\to 4}$ - - - - - $P(D_{10}|C_5)$

$\left.\begin{matrix} P_{C_1} \\ P_{C_2} \\ P_{C_3} \\ P_{C_4} \\ P_{C_5} \end{matrix}\right\}$

Ex 5Class data

D_1	D_2	D_3
3 values	4 values	2 values
R	α	Y
G	β	N
B	γ	
	δ	

Total Probab to find

- 5 Class Priori Probab
  $P_{C_1}, P_{C_2}, P_{C_3}, P_{C_4}, P_{C_5}$

5+15+20

10

50

Probab

- $P(D_1|C_i)$ 5 class x each class 3 probab
- $P(D_2|C_i)$ 5 class x " " 4 "
- $P(D_3|C_i)$ 5 class x " " 2 "

ex 5class data

D1	D2	D3
3values	4values	2values
R	α	Y
G	β	N
B	γ	
	δ	

Optimize

$4 + 5\times2 + 5\times3 + 5\times1$

$4+10+15+5 = \underline{34}$ ✓

$\boxed{5+15+20 \quad 10 \quad 50 \quad \text{Probab}}$

Total Probab to find

- 5 class Priori Probab
  $\underbrace{P_{C_1}, P_{C_2}, P_{C_3}, P_{C_4}, P_{C_5}}$ ⇒ only 4 needed
- $P(R|C_1)\,P(G|C_1)$ $\quad P(B|C_1) = 1 - P(R|C_1) - P(G|C_1)$
- $P(D_1|C_i)$ $\quad$ 5 class × each class 2 probab
- $P(D_2|C_i)$ $\quad$ 5 class × " " 3 "
- $P(D_3|C_i)$ $\quad$ 5 class × " " 1 "

Original : $K + KVD$ } Optimize $(K-1) + K(V-1)D$

Given a discrete K-class dataset containing N points, where sample points are described using D features with each feature capable of taking V values, how many parameters need to be estimated for Naïve Bayes Classifier?

(A)	$V^D K$	(C)	VDK
(B)	K^{V^D}	(D)	$K(V+D)$

$P(D_1|C_1) \to V$

$P(D_1|C_2) \to V$

$P(D_1|C_K) \to V$

D_1	D_2	D_3	Class
R	α	a	C_1
B	β	b	C_2
G	γ	c	C_3

- $P(R|C_1)+P(B|C_1)=1$ ~~X~~
-

~~a)~~ $P(R|C_1)+P(\alpha|C_1)+P(a|C_1)=1$

~~b)~~ $P(R|C_1)+P(R|C_2)+P(R|C_3)=1$

c) $P(R|C_1)+P(B|C_1)+P(G|C_1)=1$

d) $P(C_1|R)+P(C_2|R)+P(C_3|R)=1$

~~e)~~ $P(\alpha|C_1)+P(\beta|C_2)+P(\gamma|C_3)=1$

Q. C number of classes

D number of dimension

each dimension take K values.

No of Parameters for naive Bayes

a) $(C-1) + CD(K-1)$

b) $(C-1) + (C-1)DK$

c) $(C-1) + CK(D-1)$

d) all are correct

C class,

D dimension

K values in each dimension

- $(C-1)$

- $\boxed{P(D_1/C_1)}$ → K values → $(K-1)$ Probab

$\boxed{P(D_2|C_1)}$

$\boxed{P(D_3|C_1)}$

⋮ D dimension

Q1-1: Which of the following about Naive Bayes is incorrect?

H.W

- A Attributes can be nominal or numeric
- B Attributes are equally important
- C Attributes are statistically dependent of one another given the class value
- D Attributes are statistically independent of one another given the class value
- E All of above

Q1-2: Consider a classification problem with two binary features, $x_1, x_2 \in \{0,1\}$. Suppose $P(Y = y) = 1/32$, $P(x_1 = 1 | Y = y) = y/46$, $P(x_2 = 1 | Y = y) = y/62$. Which class will naive Bayes classifier produce on a test item with $x_1 = 1$ and $x_2 = 0$?

H.W

- A 16
- B 26
- C 31
- D 32

Q1-3: Consider the following dataset showing the result whether a person has passed or failed the exam based on various factors. Suppose the factors are independent to each other. We want to classify a new instance with Confident=Yes, Studied=Yes, and Sick=No.

H.W

Confident	Studied	Sick	Result
Yes	No	No	Fail
Yes	No	Yes	Pass
No	Yes	Yes	Fail
No	Yes	No	Pass
Yes	Yes	Yes	Pass

- A Pass
- B Fail

The naive Bayes classifier is used to solve a two-class classification problem with class-labels y_1, y_2. Suppose the prior probabilities are $P(y_1) = \frac{1}{3}$ and $P(y_2) = \frac{2}{3}$. Assuming a discrete feature space with

$$P(x|y_1) = \frac{3}{4} \quad \text{and} \quad P(x|y_2) = \frac{1}{4}$$

for a specific feature vector x. The probability of misclassifying x is __________

(Round off to two decimal places)

Example: avocados

You have a green-black avocado, and want to know if it is ripe. Based on this data, would you predict that your avocado is ripe or unripe?

Ripe ✓

MAP rule $P(\text{ripe} / \text{green black})$

$P(\text{unripe} \mid \text{green back})$

Train Data:

Dimension label.

color	ripeness
bright green	unripe
green-black	ripe
purple-black	ripe
green-black	unripe
purple-black	ripe
bright green	unripe
green-black	ripe
purple-black	ripe
green-black	ripe
green-black	unripe
purple-black	ripe

$P_{ripe} = \frac{7}{11}$

$P_{unripe} = \frac{4}{11}$

$P(G\text{-}B / \text{ripe}) = \frac{3}{7}$

$P(G\text{-}B \mid \text{unripe}) = \frac{2}{4}$

Test Point: green-black avocado

$$P(\text{ripe} \mid \text{green-black}) = \frac{P(\text{ripe})\,P(\text{green-black} / \text{ripe})}{P(\text{green-black})}$$

$$\frac{\frac{7}{11} \times \frac{3}{7}}{\beta} = \frac{3/11}{\beta}$$

$$P(\text{unripe} \mid \text{green-black}) = \frac{P(\text{unripe})\,P(\text{green black} \mid \text{unripe})}{P(\text{green black})}$$

$$\frac{\frac{4}{11} \times \frac{2}{4}}{\beta} = \frac{2/11}{\beta}$$

$$P(\text{Ripe} \mid GB) = \frac{3/11}{B} \Rightarrow 3/5 = .6$$

$$B = \frac{3}{11} + \frac{2}{11}$$

$$P(\text{Unripe} \mid GB) = \frac{2/11}{B} = 2/5 = .4$$

$$\text{Ripe} \rightarrow GB \leftarrow \text{unripe}$$

GB dikhai diya

* Jab bhi GB dikhega we decide Ripe hai

$P_{correct} = .6$ ✓

$P_{error} = .4$

Suppose $d = 10$, and each attribute x_i can take on 5 values. Let the number of classes be 3. Then, how many conditional probabilities do we need to estimate from the training data?

For naive Bayes.

~~a)~~ 153

~~b)~~ 122

~~c)~~ 102

~~d)~~ 137

$\boxed{(3-1) + 10 \times 3 \times 4}$

$2 + 120$

$\boxed{(b)\ 122}$

$3 + 10 \times 3 \times 5$

153

$P(\alpha/c_1) + P(\beta/c_1) + P(\gamma/c_1) + P(\delta/c_1) + \boxed{P(\sigma/c_1)} = 1$

$\boxed{(3-1) + (10 \times 3 \times (5-1))}$

Consider a Bayes classifier with 3 boolean input variables, X_1, X_2, and X_3, and one boolean output, Y.

How many parameters must be estimated to train such a ~~Bayes~~ classifier?

Naive Bayes.

a) 14
~~b)~~ 7
c) 9
d) 10

$(2-1) + 3 \times 2 \times (2-1)$

$1+6=7$

$2+3\times2\times2$

14 ✓

2 Classes.

Suppose that you are trying to solve a binary classification problem, and your data set has 4 attributes. Each attribute can take 3 possible values.

Naive Bayes

If you modeled the ~~full joint distribution~~ of the attributes and the class label, how many parameters would you need?

a) 26
b) 17
c) 23
d) 14

$$(2-1) + 4 \times 2 \times (3-1)$$

$$1 + 16 = 17$$

[2 points] Again, consider the two class problem where class label $y \in \{T, F\}$ and each training example X has 2 binary attributes X_1, X_2 $\in \{T, F\}$. How many parameters will you need to estimate if we do **not** make the Naïve Bayes conditional independence assumption?

$(2-1) + \underline{2} \times \underline{2} \times (\underline{2-1})$

(a) 3

(b) 5

$1+4=5$

(c) 7

(d) 8

age	income	student	credit	C_i: buy
youth	high	no	fair	C_2: no
youth	high	no	excellent	C_2: no
middle-aged	high	no	fair	C_1: yes
senior	medium	no	fair	C_1: yes
senior	low	yes	fair	C_1: yes
senior	low	yes	excellent	C_2: no
middle-aged	low	yes	excellent	C_1: yes
youth	medium	no	fair	C_2: no
youth	low	yes	fair	C_1: yes
senior	medium	yes	fair	C_1: yes
youth	medium	yes	excellent	C_1: yes
middle-aged	medium	no	excellent	C_1: yes
middle-aged	high	yes	fair	C_1: yes
senior	medium	no	excellent	C_2: no

Yes

$P_Y = 9/14$

$P_N = 5/14$

$P(X|Y)P(Y) = .0282$

$P(X|N)P(N) = .00685$

The sample we wish to classify is

X = (age = youth, income = medium, student = yes, credit = fair)

$P(\text{youth}|Y) = 2/9$

$P(\text{medium}|Y) = 4/9$

$P(\text{Yes}|Y) = 6/9$

$P(\text{fair}|Y) = 6/9$

$P(\text{youth}|N) = 3/5$

$P(\text{medium}|N) = 2/5$

$P(\text{yes}|N) = 1/5$

$P(\text{fair}|N) = 2/5$

Suppose we are given the following dataset, where A, B, C are input binary random variables, and y is a binary output whose value we want to predict.

$P_0 = 3/7$

$P_1 = 4/7$

$P(A=0/y=0) = 2/3$

$P(B=0/y=0) = 1/3$

$P(C=1/y=0) = 1/3$

A	B	C	y
0	0	1	0
0	1	0	0
1	1	0	0
0	0	1	1
1	1	1	1
1	0	0	1
1	1	0	1

$A=0, B=0, C=1$??

$P(X/y=0)P(y=0) = \frac{2}{3^3} \times \frac{3}{7} = .031$

$P(X|y=1)\,P(y=1) = \frac{4}{4^3} \cdot \frac{4}{7} = .035$

How would a **naive** Bayes classifier predict y given this input:

$\boxed{A = 0, B = 0, C = 1.}$ Assume that in case of a tie the classifier always prefers to predict 0 for y.

$y = 1$ ✓

$P(A=0|y=1) = 1/4$

$P(B=0|y=1) = 2/4$

$P(C=1|y=1) = 2/4$

A_1	A_2	Class Label Y
True	True	+
True	True	+
True	False	−
False	False	+
False	True	−
False	True	−
False	False	−
True	False	+

$P_+ = 1/2$

$P_- = 1/2$

$P(A_1 = \text{True} \mid +) = 3/4$

$P(A_2 = \text{True} \mid +) = 2/4$

$P(A_1 = \text{True} \mid -) = 1/4$

$P(A_2 = \text{True} \mid -) = 2/4$

$+$ class.

$P(X \mid +) P_+ = \frac{3}{4} \times \frac{2}{4} \times \frac{1}{2} \Rightarrow 3/16$

$P(X \mid -) P_- = 1/4 \times \frac{2}{4} \times \frac{1}{2} \Rightarrow 1/16$

$P(+ \mid T, T) = \frac{3/16}{\beta} \Rightarrow 3/4$

$P(- \mid T, T) = 1/16 / \beta \Rightarrow 1/4$

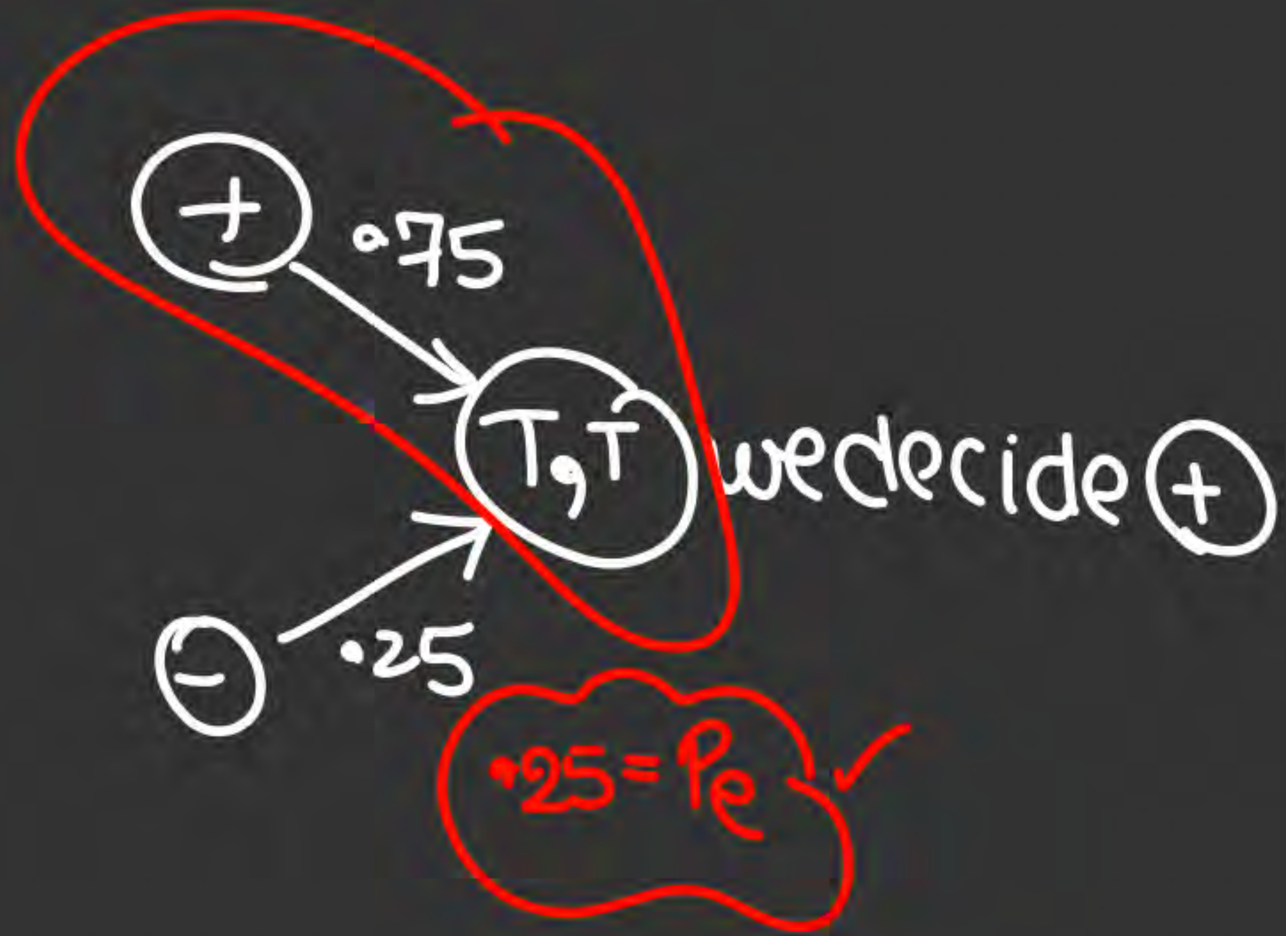
What is the probability $P(Y = + \mid A_1 = \text{True}, A_2 = \text{True})$ by applying the Naïve Bayes classifier (show each step in computing the probability)?

What label should you predict for input $(A_1 = \text{True}, A_2 = \text{True})$ based on your result above ?

$+$

$$P(+/T,T) = .75$$

$$P(-/T,T) = .25$$



Suppose you have the following training set with three boolean input x, y and z, and a boolean output U.

x	y	z	U
1	0	0	0
0	1	1	0
0	0	1	0
1	0	0	1
0	0	1	1
0	1	0	1
1	1	0	1

$P_0 = 3/7$
$P_1 = 4/7$

$P(x=0|0) = 2/3$
$P(y=1|0) = 1/3$
$P(z=0|0) = 1/3$
$P(x=0|1) = 2/4$
$P(y=1|1) = 2/4$
$P(z=0|1) = 3/4$

$x=0, y=1, z=0,$
Class ①.

Supppose you have to predict U using a naive Bayes classifier,

(a) *(3 points)* After learning is complete what would be the predicted probability $\boxed{P(U=0|x=0,y=1,z=0)?}$ $\rightarrow \boxed{.2283}$

(b) *(3 points)* Using the probabilities obtained during the Bayes Classifier training, what would be the predicted probability $\boxed{P(U=0|x=0)?}$

$P(x=0,y=1,z=0/u=0)P_{u=0}$
$\Rightarrow .0317$

$P(x=0,y=1,z=0/u=1)P_{u=1}$
$= .1071$

$P(u=0/x=0,y=1,z=0) = \frac{.0317}{\beta}$
$\hookrightarrow .2283$

$P(u=1/x=0,y=1,z=0) = \frac{.1071}{\beta}$
$\hookrightarrow .7716$

$P(u=0|x=0) = \frac{P(x=0|u=0)P_{u=0}}{P(x=0)} \Rightarrow \frac{2/7}{\beta}$

$P(u=1|x=0) = \frac{P(x=0/u=1)P_{u=1}}{P(x=0)} = \frac{2/7}{\beta}$

a) $P(u=0|x=0) + P(u=1/x=0) = 1$

b) $P(u=0|x=0) + P(u=0|x=1) = 1$

c) $P(u=1|x=0) + P(u=1|x=1) = 1$

d) $P(u=1|x=1) + P(u=0|x=1) = 1$

$$\frac{2/7}{\beta} + \frac{2/7}{\beta} = 1 \quad \beta = 4/7$$

$\Rightarrow P(u=0/x=0) = 1/2$

Consider a classification problem with 10 classes $y \in \{1, 2, \ldots, 10\}$, and two binary features $x_1, x_2 \in \{0, 1\}$. Suppose $p(Y = y) = 1/10$, $p(x_1 = 1 \mid Y = y) = y/10$, $p(x_2 = 1 \mid Y = y) = y/540$. Which class will naive Bayes classifier produce on a test item with $(x_1 = 0, x_2 = 1)$?

(A) 1 (B) 3 (C) 5 (D) 8 (E) 10

[4 points] Consider a naive Bayes classifier with 3 boolean input variables, X_1, X_2 and X_3, and one boolean output, Y.

- How many parameters must be estimated to train such a naive Bayes classifier? (you need not list them unless you wish to, just give the total)

H.W

[2 points] Suppose you have a three class problem where class label $y \in 0, 1, 2$ and each training example X has 3 binary attributes $X_1, X_2, X_3 \in 0, 1$. How many parameters do you need to know to classify an example using the Naive Bayes classifier?

(a) 5

(b) 9

(c) 11

(d) 13

(e) 23

$(3-1) + 3 \times 3 \times (2-1)$

$2 + 9 \times 1$

$= 11$

Consider a Naïve Bayes classifier with $\boxed{\text{100 feature dimensions.}}$ The label is binary with $\boxed{P(y = 0) = P(y = 1) = 0.5}$. All features are binary and have the same conditional probabilities:

$$\boxed{P(x_i = 1 \mid y = 0) = a}$$

$$\boxed{P(x_i = 1 \mid y = 1) = b}$$

$P(x_2=1|y=1)=b$

for $i = 1, \ldots, 100$. The values of a and b are within $[0, 1]$.

Test.

Given an item with alternating feature values $\mathbf{X} = (x_1 = 1, x_2 = 0, x_3 = 1, \ldots, x_{100} = 0)$, compute $P(y = 1 \mid \mathbf{X})$.

$$\frac{P(x/y=1)\,P_{y=1}}{P(x)} \qquad P_{y=1} \to .5$$

$$P(x/y=1) \to [\underbrace{P(x_1=1/y=1)}_{b}\ \underbrace{P(x_2=0|y=1)}_{(1-b)}\ \underbrace{P(x_3=1|y=1)}_{b}\ ---]$$

$$b^{50}(1-b)^{50}$$

$$P(y=1|x) = \frac{.5\, b^{50}(1-b)^{50}}{P(x)}$$

$$P(y=0|x) = \frac{P_{y=0}\, P(x|y=0)}{P(x)} = \frac{.5\, a^{50}(1-a)^{50}}{P(x)}$$

$$P(y=1|x) = \left[\frac{.5\, b^{50}(1-b)^{50}}{.5\, b^{50}(1-b)^{50} + .5\, a^{50}(1-a)^{50}}\right] = \frac{b^{50}(1-b)^{50}}{a^{50}(1-a)^{50} + b^{50}(1-b)^{50}}$$

THANK - YOU