Data Science and Artificial Intelligence

Machine Learning

Bayesian Learning

Lecture No. 4

By- SIDDHARTH SABHARWAL SIR

Recap of Previous Lecture

- Topic — Naive Bayes.
- Topic
- Topic
- Topic
- Topic

Topics to be Covered

Topic — Naive Bayes

Topic — Questions.

Topic

Topic

Topic

ONE SMALL

POSITIVE THOUGHT

IN THE

MORNING

CAN CHANGE YOUR

WHOLE DAY

Laplace Smoothening

$\boxed{\text{we use } \alpha}$

Zeroprobabilities ko

exist due to lack of data

⇓

- nonzero banane keliye we do laplace smoothing

Bayesian Decision Theory

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

done

$$\langle Outlook = sunny, Temperature = cool, Humidity = high, Wind = strong \rangle$$

Bayesian Decision Theory

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	85	85	False	No
D2	Sunny	80	90	True	No
D3	Overcast	83	86	False	Yes
D4	Rainy	70	96	False	Yes
D5	Rainy	68	80	False	Yes
D6	Rainy	65	70	True	No
D7	Overcast	64	65	True	Yes
D8	Sunny	72	95	False	No
D9	Sunny	69	70	False	Yes
D10	Rainy	75	80	False	Yes
D11	Sunny	75	70	True	Yes
D12	Overcast	72	90	True	Yes
D13	Overcast	81	75	False	Yes
D14	Rainy	71	91	True	No

P(Humidity|Yes)

$$\boxed{\begin{array}{c}86\\96\\80\\65\\70\\80\\70\\90\\75\end{array}}$$

$\rightarrow \mu$

$\rightarrow \sigma^2$

$$\rightarrow \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$$

Bayesian Decision Theory

Color	Type	Origin	Stolen?
Red	Sports	Domestic	Yes
Red	Sports	Domestic	No
Red	Sports	Domestic	Yes
Yellow	Sports	Domestic	No
Yellow	Sports	Imported	Yes
Yellow	SUV	Imported	No
Yellow	SUV	Imported	Yes
Yellow	SUV	Domestic	No
Red	SUV	Imported	No
Red	Sports	Imported	Yes
Red	SUV	Domestic	?

- $P_Y = 1/2$
- $P_N = 1/2$

Room की updating Probab

Color $P(R/Y) = \frac{3}{5}$ $P(\text{yellow}/Y) = 2/5$

$P(R/N) = \frac{2}{5}$ $P(\text{yellow}/N) = \frac{3}{5}$

$\boxed{\text{Laplace smoothing}(\alpha)}$

Yes Class Point

R
R
R
Yellow
Yellow

$\rightarrow$

R, R, R
α points of Red
Ye, Ye
α points of Ye

$P(R/Y) = \frac{3+\alpha}{5+2\alpha}$

$P(\text{Yellow}/Y) = \frac{2+\alpha}{5+2\alpha}$

Bayesian Decision Theory

Color	Type	Origin	Stolen?
Red	Sports	Domestic	Yes
Red	Sports	Domestic	No
Red	Sports	Domestic	Yes
Yellow	Sports	Domestic	No
Yellow	Sports	Imported	Yes
Yellow	SUV	Imported	No
Yellow	SUV	Imported	Yes
Yellow	SUV	Domestic	No
Red	SUV	Imported	No
Red	Sports	Imported	Yes
Red	SUV	Domestic	?

- find class of Red, SUV, domestic.

$P[Yes/Red, SUV, domestic]$

and

$P(No/Red, SUV, domestic)$

$\boxed{P(Y/X)}$ Posterior, Bigger decide class.

$P(R, SUV, D/Yes)P(Yes)$ $\quad$ $P(R, SUV, D/No)P(No)$

$P(R/Yes)P(SUV/Yes)P(D/Yes)P_{Yes}$

$\left[\frac{3}{5} \times \frac{1}{5} \times \frac{2}{5} \times \frac{1}{2}\right] \Rightarrow \frac{3}{125}$

Bayesian Decision Theory

Color	Type	Origin	Stolen?
Red	Sports	Domestic	Yes
Red	Sports	Domestic	No
Red	Sports	Domestic	Yes
Yellow	Sports	Domestic	No
Yellow	Sports	Imported	Yes
Yellow	SUV	Imported	No
Yellow	SUV	Imported	Yes
Yellow	SUV	Domestic	No
Red	SUV	Imported	No
Red	Sports	Imported	Yes
Red	SUV	Domestic	?

No

• find class of Red, SUV, domestic

P[Yes/Red, SUV, domestic]

and

P(No, /Red, SUV, domestic)

$\boxed{P(Y/x)}$ Posterior, Bigger decide class.

$P(R, S_u, D/Yes)P(Yes)$ $P(R, S_u, D/No)P(No)$

$P(R/No)P(Su/No)P(D/No)P(No)$

$\left(\frac{2}{5} \times \frac{3}{5} \times \frac{3}{5} \times \frac{1}{2}\right) = \frac{9}{125}.$

- $P(R/No) + P(R/Y) = 1$ False.

Colour has 2 option only Red, Yellow

$$P(\overline{Red}/\underline{Y}) + P(\overline{Yellow}|\underline{Y}) = 1$$

$$P(Red|No) + P(Yellow|No) = 1$$

??

$$P(Yes/R, Su, D) = \frac{P(R, Su, D/Yes)\,P(yes)}{P(R, Su, D)} \Rightarrow \frac{3}{125}$$

$$P(No/R, Su, D) = \frac{P(R, Su, D/No)\,P(No)}{P(R, Su, D)} \rightarrow \frac{9}{125}$$

$$P_1 = \frac{3/125}{\beta}$$

$$P_2 = \frac{9/125}{\beta}$$

$$P_1 + P_2 = 1$$

$$\frac{3/125}{\beta} + \frac{9/125}{\beta} = 1$$

$$\beta = 12/125$$

$$P_1 = 1/4 \checkmark$$

$$P_2 = 3/4 \checkmark$$

Naive Bayes: Assumption

- dimension are independent for a given clan.

$$P(R, Su, D) = P(R)\,P(Su)\,P(Do)$$

$$= \frac{\text{No. of Red}}{\text{Total No of points}} \times \frac{\text{No of Suv}}{\text{Total No of}} \times \frac{\text{No of Domestic}}{\text{Total No of points}}$$

$$P(R, Su, Do/Yes) = P(R/Yes)\,P(Su/Yes)\,P(Do/Yes)$$

$$\text{New probab} \Rightarrow \frac{\text{Old value} + \alpha}{\text{Old value} + K\alpha}$$

K is No of values the dimension Can take

α

0 → V.large

Overfit
Zero probab
Probab

Underfit

- α is hyperparameter
- Best value of α, use CV.

Bayesian Decision Theory

department	*status*	*age*	*salary*
sales	senior	31...35	46K...50K
sales	junior	26...30	26K...30K
sales	junior	31...35	31K...35K
systems	junior	21...25	46K...50K
systems	senior	31...35	66K...70K
systems	junior	26...30	46K...50K
systems	senior	41...45	66K...70K
marketing	senior	36...40	46K...50K
marketing	junior	31...35	41K...45K
secretary	senior	46...50	36K...40K
secretary	junior	26...30	26K...30K

Naïve Bayes Algorithm

Laplace smoothing

Naïve Bayes Algorithm

What if the dimension are continuous in nature

The numeric weather data with summary statistics

outlook			temperature		humidity		windy			play	
	yes	no	yes	no	yes	no		yes	no	yes	no
sunny	2	3	83	85	86	85	false	6	2	9	5
overcast	4	0	70	80	96	90	true	3	3		
rainy	3	2	68	65	80	70					
			64	72	65	95					
			69	71	70	91					
			75		80						
			75		70						
			72		90						
			81		75						

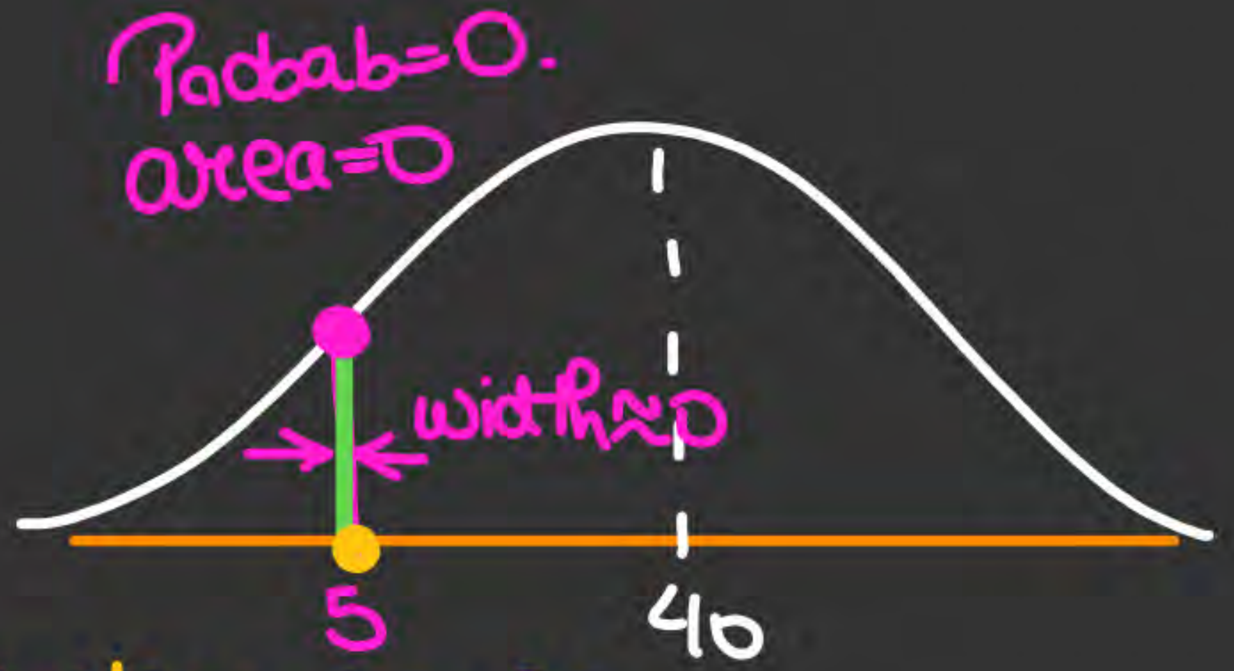
Bayesian Decision Theory

The numeric weather data with summary statistics											
outlook			temperature		humidity		windy			play	
	yes	no	yes	no	yes	no		yes	no	yes	no
sunny	2	3	83	85	86	85	false	6	2	9	5
overcast	4	0	70	80	96	90	true	3	3		
rainy	3	2	68	65	80	70					
			64	72	65	95					
			69	71	70	91					
			75		80						
			75		70						
			72		90						
			81		75						

Q $X = N(40, 100)$

$P(X=5) = $ Zero.

$P(X=5) \not\propto$ value of PDF @ $X=5$



* PDF ke integration se Probab aati hai

Bayesian Decision Theory

find class $W=168$, foot size $= 9$.

$P_M = P_F = .5$

• $P(168/M)P(9/M)P_M$

and

• $P(168/F)P(9/F)P_F$

Person (label)	[redacted]	Weight (lbs)	Foot size (inches)
Male	[redacted]	180	12
Male	[redacted]	190	11
Male	[redacted]	170	12
Male	[redacted]	165	10
Female	[redacted]	100	6
Female	[redacted]	150	8
Female	[redacted]	130	7
Female	[redacted]	150	9

Male weight: $\mu = 176.25$, $\sigma^2 = 92.18$

Male foot size: $\mu = 11.25$, $\sigma^2 = .6875$

Female weight: $\sigma^2 = 418.75$, $\mu = 132.5$

Female foot size: $\mu = 7.5$, $\sigma^2 = 1.25$

$P(168/F)P(9/F)P_F$

$$\frac{1}{\sqrt{2\pi 418.75}} e^{-\frac{(168-132.5)^2}{2\times 418.75}} \times \frac{1}{\sqrt{2\pi \times 132.5}}$$

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$$

$$\boxed{P(168|M) = \frac{1}{\sqrt{2\pi \times 92.18}} e^{-(168-176.25)^2/2\times 92.18}}$$

Bayesian Decision Theory

Find class Male ✓
$W = 168$, Foot size $= 9$.

$P_M = P_F = .5$

$P(168/M)\,P(9/M)\,P_M$

$$\frac{1}{\sqrt{2\pi\, 92.18}}\, e^{-\frac{(168-176.25)^2}{2\times 92.18}} \cdot \frac{1}{\sqrt{2\pi\, .6875}}\, e^{-\frac{(11.25-9)^2}{2\times .6875}} \times \frac{1}{2}$$

$= 1.212\times10^{-3}$

label

$P(168/F)\,P(9/F)\,P_F$

$$\frac{1}{\sqrt{2\pi\, 418.75}}\, e^{-\frac{(168-132.5)^2}{2\times 418.75}} \times \frac{1}{\sqrt{2\pi\times 1.25}}\, e^{-\frac{(9-7.5)^2}{2\times 1.25}} \times \frac{1}{2}$$

3.14×10^{-4} ✓

Person	Weight (lbs)	Foot size (inches)
Male	180	12
Male	190	11
Male	170	12
Male	165	10
Female	100	6
Female	150	8
Female	130	7
Female	150	9

Male weight: $\mu = 176.25$, $\sigma^2 = 92.18$

Male foot size: $\mu = 11.25$, $\sigma^2 = .6875$

Female weight: $\sigma^2 = 418.75$, $\mu = 132.5$

Female foot size: $\mu = 7.5$, $\sigma^2 = 1.25$

age	Education	Employed (label)
40	G	Y
50	G	Y
60	PG	Y
30	PG	N
25	PG	N
55	G	N

$P(Age/Y) = \checkmark$

$N(50, 66.6)$

$P(Age/N) = \checkmark$

$N(36.66, 172.2)$

Education

$P(G/Y) = 2/3$

$P(PG/Y) = 1/3$

$P(G/N) = 1/3$

$P(PG/N) = 2/3.$

Test

45, PG

$P(Age/Y) =$ ✓

$N(50, 66.6)$

$P(Age/N) =$ ✓

$N(36.66, 172.2)$

$P_Y = P_N = 1/2$

NO. ✓

$\bullet\ P(45/Y)\,P(PG/Y)\,P_Y$

$$\frac{1}{\sqrt{2\pi \times 66.6}}\, e^{-(45-50)^2/2\times 66.6} \times \frac{1}{3} \times \frac{1}{2}$$

6.75×10^{-3}

$P(45/N)\,P(PG/N)\,P_N$

$$\frac{1}{\sqrt{2\pi\, 172.2}}\, e^{-(36.66-45)^2/2\times 172.2} \times \frac{\cancel{2}}{3} \times \frac{1}{\cancel{2}}$$

$\boxed{8.25 \times 10^{-3}}$

Education

$P(G/Y) = 2/3$

$P(PG/Y) = 1/3$

$P(G/N) = 1/3$

$P(PG/N) = 2/3$

Test

45, PG

Q: Consider a classification problem with 10 classes $y \in \{1,2,\ldots,10\}$, and two binary features $x1, x2 \in \{0,1\}$.

Suppose:

$p(Y=y)=1/10$,

$p(x1=1|Y=y)=y/10$,

$p(x2=1|Y=y)=y/540$

$P(X_1=1/Y=y)=y/10$

$P(X_1=0/Y=y)=1-\frac{y}{10}$

Which class will naïve Bayes classifier produce on a test item with $(x1=0,x2=1)$?

A. 1

B. 3

C. 5 ✓

D. 8

E. 10

So we want to find the 'y' which max

$P(X_1=0, X_2=1/Y=y)$

$\Downarrow$

$\underbrace{P(X_1=0|Y=y)\,P(X_2=1|Y=y)\,P(Y=y)}$

Value of 'y' that max this equation is ans.

$\left(1-\frac{y}{10}\right)\frac{y}{540}\,\frac{1}{10}$

$=\left(\frac{y}{540}-\frac{y^2}{5400}\right)\frac{1}{10}$,

$$\frac{d}{dy}\left(\frac{y}{540}-\frac{y^2}{5400}\right)\times\frac{1}{10}=0$$

$$\frac{1}{540}-\frac{2y}{5400}=0$$

$$y=5$$ ✓

- MAP Rule

$\Downarrow$

- $P(y_1/x) \underset{y_2}{\overset{y_1}{\gtrless}} P(y_2/x)$

example ✓

$P(x|y_1) = .8$

$P(x|y_2) = .5$

$P_{y_1} = .4$

$P_{y_2} = .6.$

$\boxed{P_{y_1} P(x|y_1)} \gtrless P_{y_2} P(x|y_2)$

$.4 \times .8 \gtrless .6 \times .5$

$.32 \gtrless .30$

$\Downarrow$

$P_{y_1} P(x|y_1) = .32$

$P_{y_2} P(x|y_2) = .30.$

If x is given then we decide y_1.

- MAP Rule

$\Downarrow$

- $P(y_1/x) \overset{y_1}{\underset{y_2}{\gtrless}} P(y_2/x)$

$\boxed{P_{y_1} P(x/y_1)} \gtrless P_{y_2} P(x/y_2)$

$.4 \times .8 \gtrless .6 \times .5$

$.32 \gtrless .30$

$\Downarrow$

$P_{y_1} P(x/y_1) = .32$

$P_{y_2} P(x/y_2) = .30.$

example ✓

$P(x/y_1) = .8 \qquad P_{y_1} = .4$

$P(x/y_2) = .5 \qquad P_{y_2} = .6.$

If x is given then we decide y_1.

$.516 \leftarrow P(y_1/x) = \dfrac{P(y_1) P(x/y_1)}{P_x} = \dfrac{.32}{\beta}$

$.483 \leftarrow P(y_2/x) = \dfrac{P_{y_2} P(x/y_2)}{P_x} = \dfrac{.30}{\beta}$

$\dfrac{.32}{\beta} + \dfrac{.30}{\beta} = 1$

$\beta = .32 + .30$

1000 Call

1000 Call mein lo Sunai diya

Jab bhi lo sunai dega ⇒ Hello bala

But 516 Hello ✓

484 Chalo ✓

Mujhe x mila

$P(y_1|x) = .516$

$P(y_2|x) = .483$

MAP rule

$P(y_1|x) > P(y_2|x)$

.516 ✓

Hello

Call

Chalo

.483

Sunai diya

"lo"

Probab of Correct detection .516

Probab of error. .484.

Correct 516 times out of 1000

Error 484 " " " "

$P(\text{Hello}/\text{lo}) = .37$

$\boxed{P(\text{cholo}|\text{lo}) = .45}$

$\boxed{P(\text{lelo}|\text{lo}) = .18}$

$\boxed{\begin{matrix} .37 \\ \text{Hello} \end{matrix}}$

.45
Chalo

$\boxed{\begin{matrix} .18 \\ \text{lelo} \end{matrix}}$

→ lo Sunai diya

Jab bhi lo Sunai diya, decide chalo.

$P_{error} = .37 + .18$

$P_{correct} = .45$

greatest

$P(y_i|x)$

decide y_i for x always

→ decide y_4 for x always.

$P_c = .7$

$P_e = .3$

1. What type of algorithm is Naive Bayes used for in machine learning?

a. Classification

b. Regression

c. Clustering

d. Reinforcement learning

H.W

H.W

3. What is the "naive" assumption in Naive Bayes?

a. It assumes that all features are equally important.

b. It assumes that features are independent of each other.

c. It assumes that the dataset is small.

d. It assumes that features are dependent on each other.

6. In a binary classification problem, if the probability of an event occurring in Class A is 0.8 and in Class B is 0.3, what is the odds ratio in favor of Class A?

a. 0.375

H.W

b. 1.5

c. 2.67

d. 3.33

9. In the context of Naive Bayes, what is Laplace smoothing (additive smoothing) used for?

a. Reducing the impact of rare features

b. Increasing the model's complexity

c. Decreasing the training time

d. Ignoring missing data H.W

13. In a binary classification problem, a Naive Bayes classifier correctly classifies 85% of Class A instances and 90% of Class B instances. If the prior probabilities are P(Class A) = 0.4 and P(Class B) = 0.6, what is the overall accuracy of the classifier?

a. 0.48

b. 0.87

c. 0.90

d. 0.84

Total probab

$$P(Acc) = P(A)\,P(Acc|A) + P(B)\,P(Acc|B)$$

$$= .4 \times .85 + .6 \times .9$$

$$= .88$$

A Naïve Bayes text classifier is trained on a dataset with two classes: **Spam (S)** and **Not Spam (NS)**. The probabilities of the classes are:

- P(S) = 0.3
- P(NS) = 0.7

Given that a message contains the word "offer," the likelihood values are:

- P("offer" | S) = 0.8 H.W
- P("offer" | NS) = 0.2

Using Naïve Bayes, what is the probability that the message belongs to the **Spam** category, given it contains "offer"?

- (A) 0.6
- (B) 0.5
- (C) 0.8
- (D) 0.3

Suppose you are using Naïve Bayes for spam detection. You have a training dataset with the following word counts:

Word	Count in Spam (S)	Count in Not Spam (NS)
"money"	3	1
"win"	5	2
"lottery"	4	0

H.W

Total word occurrences:

- **Spam**: 30 words
- **Not Spam**: 20 words

Using **Laplace smoothing** ($\alpha = 1$), calculate **P("lottery" | Spam)**.

Naïve Bayes Classifier

How low alpha effect bias and variance..

Naïve Bayes Classifier

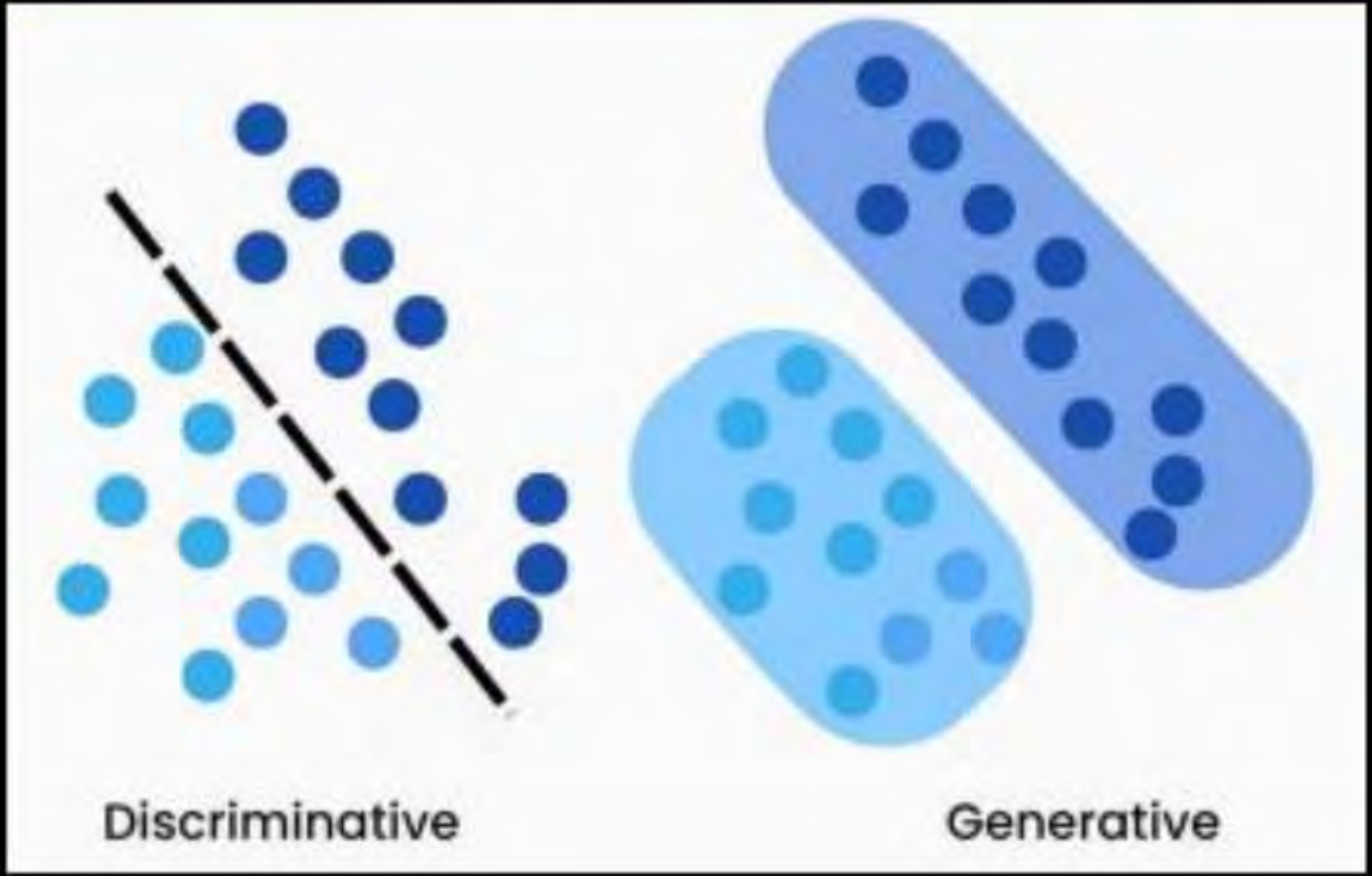
Advantages of Naïve Bayes Classifier:

- 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



Story...

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.

Discriminative vs. Generative Learning

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

There are two ways you can classify languages.

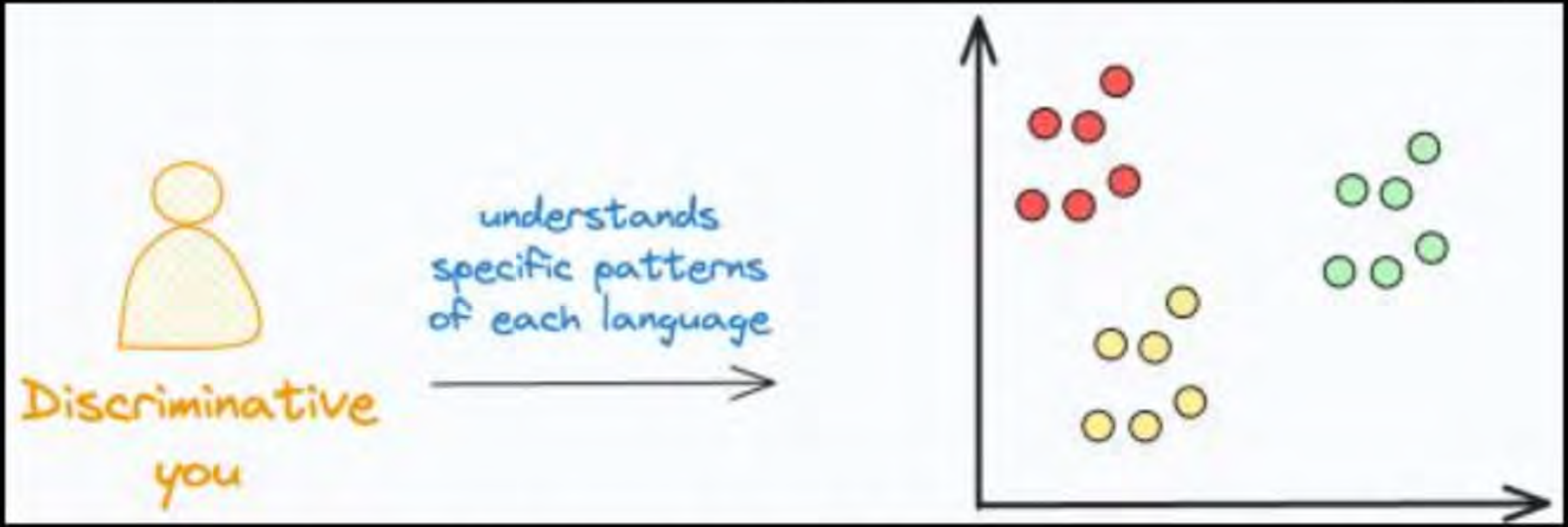
- ❑ Learn every language and then classify a new language based on acquired knowledge.
- ❑ Understand some distinctive patterns in each language without truly learning the language. Once done, classify a new language.

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.

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.

Bayesian Decision Theory

Generative Learning

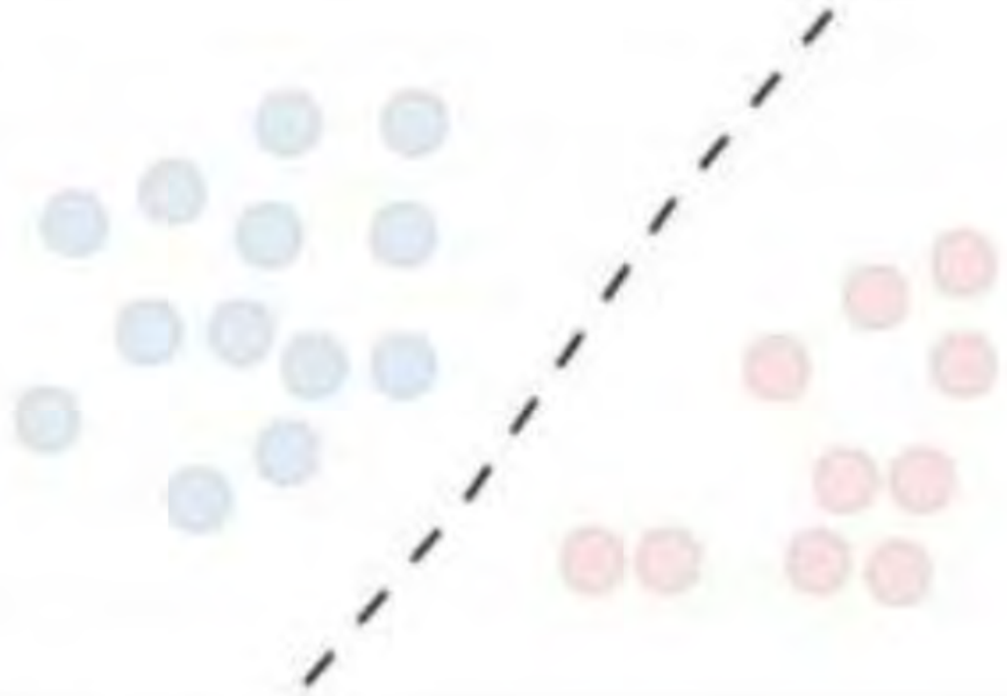
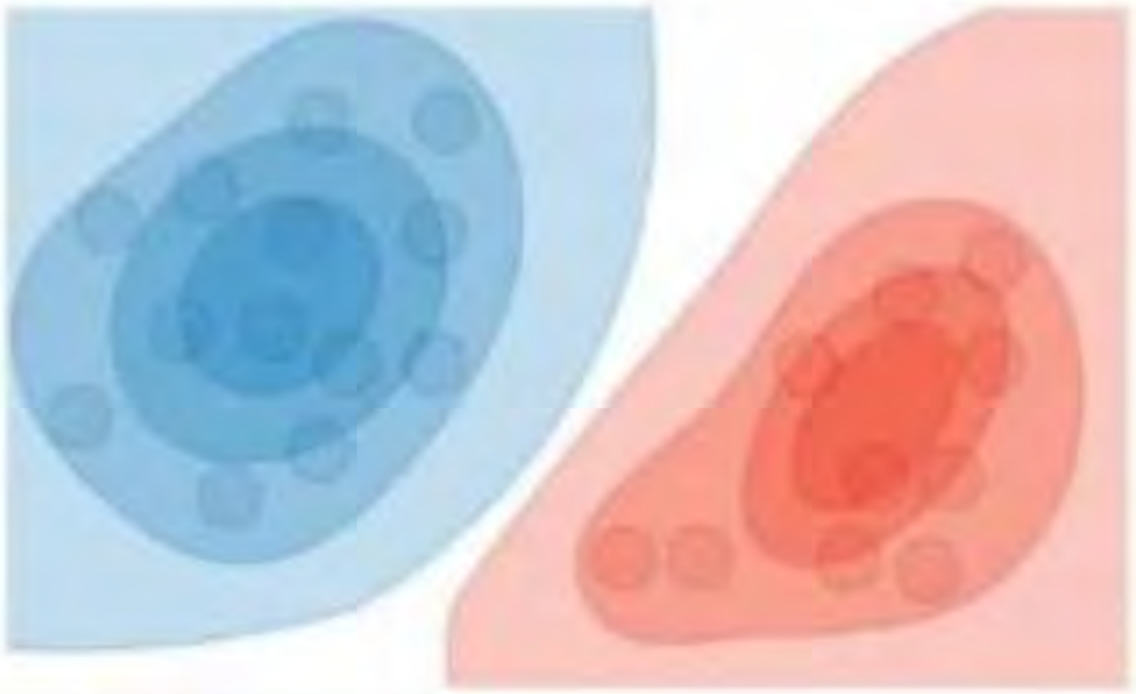
- Generative models are machine learning models that learn to generate new data samples similar to the training data they were trained on. They capture the underlying distribution of the data and can produce novel instances.
- So, the Generative approach focuses on the distribution of individual classes 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.

Bayesian Decision Theory

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		
Examples	Regressions, SVMs	GDA, Naive Bayes

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)	$K(V+D)$

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

- 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 $x_2 = 0$?

- 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.

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

$P_{y_1} = 1/3$ $\Big\}$ $P(x/y_1) = 3/4$

$P_{y_2} = 2/3$ $\Big\}$ $P(x/y_2) = 1/4$.

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)

a) .4

b) .6

c) .8

d) .2

$$\left.\begin{array}{l} P_{y_1} = 1/3 \\ P_{y_2} = 2/3 \end{array}\right\} \begin{array}{l} P(x|y_1) = 3/4 \\ P(x|y_2) = 1/4 . \end{array}$$

$$P(y_1|x) = \frac{P(y_1)P(x|y_1)}{P(x)} = \frac{1/4}{\beta}$$

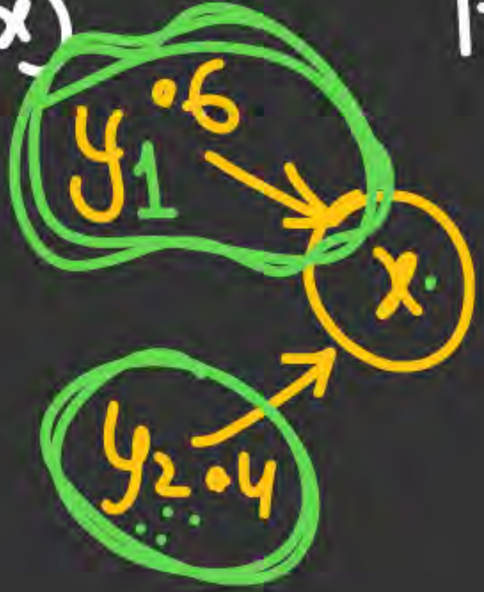
$$P(y_2|x) = \frac{P(y_2)P(x|y_2)}{P(x)} = \frac{1/6}{\beta}$$

$$\frac{\frac{1}{4}}{\beta} + \frac{1/6}{\beta} = 1$$

$$\beta = 1/4 + 1/6 = 10/24$$

$$P_{y_1|x} = .6$$

$$P_{y_2|x} = .4$$



THANK - YOU