UNIT – V PROBABILISTIC MODELS

- Conditional Probability, Joint Probability, Probability Density Function, Normal Distribution and its Geometric Interpretation, Naïve Bayes Classifier, Discriminative Learning with Maximum Likelihood. Probabilistic Models with Hidden variables: Expectation-Maximization methods, Gaussian Mixtures

Example 1

Example 1: You draw one card from a deck of cards. What's the probability that you draw an ace?

$$P(\text{draw an ace}) = \frac{\#\text{ of aces in the deck}}{\#\text{ of cards in the deck}} = \frac{4}{52} = .0769$$

Independent Events

If the probability of the occurrence of event A is the same regardless of whether or not an outcome B occurs, then the outcomes A and B are said to be independent of one another. Symbolically, if

$$P(A \mid B) = P(A)$$

then A and B are independent events.

Independent Events

$$P(A \cap B) = P(A \mid B)P(B)$$

then we can also state the following relationship for independent events:

$$P(A \cap B) = P(A)P(B)$$

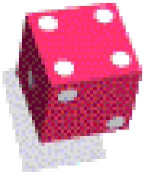
if and only if

A and B are independent events.

- Joint probability

It is the probability of occurrence of two or more events together

Example



- A coin is tossed and a single 6-sided die is rolled. Find the probability of getting a head on the coin and a 3 on the die.
- Probabilities:

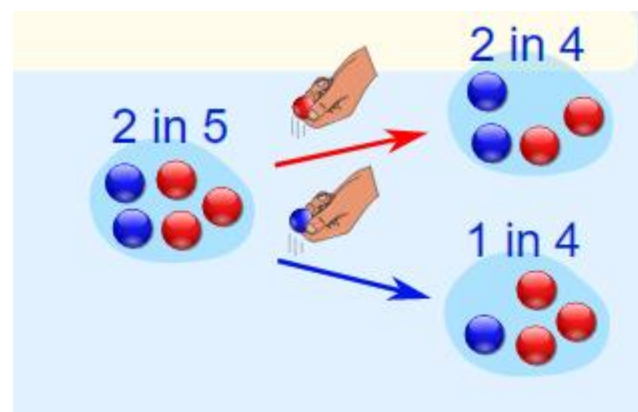
P(head) = 1/2

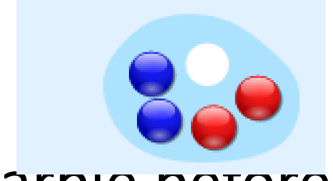
P(3) = 1/6

P(head and 3) = 1/2 * 1/6 = 1/12

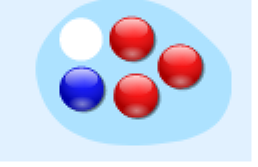
Dependent Events

- they **can be affected by previous events** ...
- Example: Marbles in a Bag
- 2 blue and 3 red marbles are in a bag.
- What are the chances of getting a blue marble?
- The chance is **2 in 5**
- **But after taking one out** the chances change!
- So the next time:
-





- if we got a **red** marble before, then the chance of a blue marble next is **2 in 4**



- if we got a **blue** marble before, then the chance of a blue marble next is **1 in 4**
- This is because we are **removing** marbles from the bag.
- So the next event **depends on** what happened in the previous event, and is called **dependent**.

- **With** Replacement: the events are **Independent** (the chances don't change)
- **Without** Replacement: the events are **Dependent** (the chances change)

- In our marbles example
- Event A is "get a Blue Marble first" with a probability of 2/5:
- **P(A) = 2/5**
- And Event B is "get a Blue Marble second" ... but for that we have 2 choices:
- If we got a **Blue Marble first** the chance is now **1/4**
- If we got a **Red Marble first** the chance is now **2/4**
- So we have to say **which one we want**, and use the symbol "|" to mean "given":

Conditional Probability

- P(B|A) means "Event B **given** Event A"
- In other words, event A has already happened, now what is the chance of event B?
- P(B|A) is also called the "Conditional Probability" of B given A.

"Probability Of" *"Given"*

$$P(A \text{ and } B) = P(A) \times P(B \mid A)$$

Event A *Event B*

$$P(B \mid A) = \frac{P(A \text{ and } B)}{P(A)}$$

- A pair of number cubes is rolled . What is the probability that both numbers are odd if their sum is 6?

  - Let *A* be the event "Both numbers are odd."
  - Let *B* be the event "The sum of the numbers is 6."
  - You need to find the probability of *A* given *B*.
  - That is, you need to find

$$P(A|B).$$

$$P\left(A \cap B\right) = \frac{\text{number of outcomes in } A \text{ and } B}{\text{number of outcomes in sample space}} = \frac{3}{36}$$

$$P\left(B\right) = \frac{\text{number of outcomes in } B}{\text{number of outcomes in sample space}} = \frac{5}{36}$$

$$P\left(A \middle| B\right) = \frac{P\left(A \cap B\right)}{P\left(B\right)} \qquad P\left(A \middle| B\right) = \frac{3}{5}$$

- If dice are rolled .What is the probability that the sum of faces will not exceed 7 ? Given that at least one face should show 4 .

- If dice are rolled .What is the probability that the sum of faces will not exceed 7 ? Given that at least one face should show 4 .

- Let A be the event that sum will not exceed 7
- Let B be the even that one face is 4
- To calculate P$(A|B)$

- P$(A|B)$=$\frac{P(A\cap B)}{P(B)}$= $\frac{11/36}{6/36}$=6/11

- In a box there are 8 red 7 blue and 6 green balls. One ball is picked up randomly .What is the probability that it is neither green nor red?

- A vacation resort offers bicycles and personal watercrafts for rent. The resort's manager made the following notes about rentals:
- 200 customers rented items in all—100 rented bicycles and 100 rented personal watercrafts.
- Of the personal watercraft customers, 75 customers were young (30 years old or younger) and 25 customers were older (31 years old or older).
- 125 of the 200 customers were age 30 or younger. 50 of these customers rented bicycles, and 75 of them rented personal watercrafts.
- Consider the following events that apply to a random customer.
  - *Y*: The customer is young (30 years old or younger).
  - *W*: The customer rents a personal watercraft.

Calculate

$\mathrm{P}(Y|W)$ $\mathrm{P}(W|Y)$

$$P(Y|W) = \frac{75}{100} = 0.75$$

$$P(W|Y) = \frac{75}{125} = 0.6$$

Bayes’ Theorem or Bayes’ Rule

- Important Theorem associated with Conditional probability
- Allows you to find P(A|B) from P(B|A), i.e. to ‘invert’ conditional probabilities.

- $P(A|B) = \dfrac{P(B|A) \cdot P(A)}{P(B)}$

- Statement:Let E_1, E_2,…,E_n be a set of events associated with a sample space S, where all the events E_1, E_2,…,E_n have nonzero probability of occurrence and they form a partition of S. Let A be any event associated with S, then according to Bayes theorem,

$$P(E_i \mid A) = \frac{P(E_i)P(A \mid E_i)}{\sum_{k=1}^{n} P(E_k)P(A|E_k)}$$

for any k = 1, 2, 3, …., n

- Using Conditional Probability formula

$$P(E_i \mid A) = \frac{P(E_i \cap A)}{P(A)} \cdots\cdots\cdots\cdots (1)$$

- Using multiplication rule of probability,

$$P(E_i \cap A) = P(E_i)P(A \mid E_i) \cdots\cdots\cdots\cdots (2)$$

- Using total probability theorem,

$$P(A) = \sum_{k=1}^{n} P(E_k)P(A|E_k) \cdots\cdots\cdots\cdots (3)$$

- Putting the values from equations (2) and (3) in

$$P(E_i \mid A) = \frac{P(E_i)P(A \mid E_i)}{\sum_{k=1}^{n} P(E_k)P(A|E_k)}$$

- A bag I contain 4 white and 6 black balls while another Bag II contains 4 white and 3

Let E_1 be the event of choosing the bag I, E_2 the event of choosing the bag II, and A be the event of drawing a black ball.

Then, $P(E_1) = P(E_2) = \frac{1}{2}$

Also, $P(A|E_1) = P$(drawing a black ball from Bag I) = $\frac{6}{10} = \frac{3}{5}$

$P(A|E_2) = P$(drawing a black ball from Bag II) = $\frac{3}{7}$

By using Bayes' theorem, the probability of drawing a black ball from bag I out of two bags,

$$P(E_1|A) = \frac{P(E_1)P(A|E_1)}{P(E_1)P(A|E_1)+P(E_2)P(A|E_2)}$$

$$= \frac{\frac{1}{2} \times \frac{3}{5}}{\frac{1}{2} \times \frac{3}{7} + \frac{1}{2} \times \frac{3}{5}} = \frac{7}{12}$$

- A man is known to speak truth 2 out of 3 times. He throws a die and reports that number obtained is a four. Find the

Let A be the event that the man reports that number four is obtained.

Let E_1 be the event that four is obtained and E_2 be its complementary event.

Then, $P(E_1)$ = Probability that four occurs = $\frac{1}{6}$

$P(E_2)$ = Probability that four does not occurs = $1 - P(E_1) = 1 - \frac{1}{6} = \frac{5}{6}$

Also, $P(A|E_1)$ = Probability that man reports four and it is actually a four = $\frac{2}{3}$

$P(A|E_2)$ = Probability that man reports four and it is not a four = $\frac{1}{3}$

By using Bayes' theorem, probability that number obtained is actually a four,

$$P(E_1|A) = \frac{P(E_1)P(A|E_1)}{P(E_1)P(A|E_1) + P(E_2)P(A|E_2)} = \frac{\frac{1}{6} \times \frac{2}{3}}{\frac{1}{6} \times \frac{2}{3} + \frac{5}{6} \times \frac{1}{3}} = \frac{2}{7}$$

- If a single card is drawn from a standard deck of playing cards. If the card is the face card what is the probability that it is a king.

I have three bags that each contain 100 marbles:
Bag 1 has 75 red and 25 blue marbles; Bag 2 has 60 red and 40 blue marbles; Bag 3 has 45 red and 55 blue marbles. I choose one of the bags at random and then pick a marble from the chosen bag, also at random. What is the probability that the chosen marble is red?

- At a certain university, 4% of men are over 6 feet tall and 1% of women are over 6 feet tall. The total student population is divided in the ratio 3:2 in favour of women. If a student is selected at random from among all those over six feet tall, what is the probability that the student is a woman?

- A factory production line is manufacturing bolts using three machines, A, B and C. Of the total output, machine A is responsible for 25%, machine B for 35% and machine C for the rest. It is known from previous experience with the machines that 5% of the output from machine A is defective, 4% from machine B and 2% from machine C. A bolt is chosen at random from the production line and found to be defective. What is the probability that it came from
- (a) machine A (b) machine B (c) machine C?

Probabilistic Models

- Discriminative models
  - Learns conditional Probability
  - A Discriminative model models the **decision boundary between the classes**.
  - Logistic regression
  - Scalar Vector Machine

- Generative models
  - Learns the joint probability
  - A Generative Model explicitly models the **actual distribution of each class**
  - Naïve Bayes

Naïve Bayes Classifier

- Generative models
  - Learns the joint probability
  - A Generative Model explicitly models the **actual distribution of each class**

Weather	Car	Class
sunny	working	go-out
rainy	broken	go-out
sunny	working	go-out
sunny	working	go-out
sunny	working	go-out
rainy	broken	stay-home
rainy	broken	stay-home
sunny	working	stay-home
sunny	broken	stay-home
rainy	broken	stay-home

- Convert this into numbers. :
- **Variable: Weather**
- sunny = 1
- rainy = 0
- **Variable: Car**
- working = 1
- broken = 0
- **Variable: Class**
- go-out = 1
- stay-home = 0

Weather	Car	Class
1	1	1
0	0	1
1	1	1
1	1	1
1	1	1
0	0	0
0	0	0
1	1	0
1	0	0
0	0	0

- There are two types of quantities that need to be calculated from the dataset
  - Class Probabilities.
  - Conditional Probabilities.
- **Calculate the Class Probabilities**
- The dataset is a two class problem a
- P(class=1) = count(class=1) / (count(class=0) + count(class=1))
- P(class=0) = count(class=0) / (count(class=0) + count(class=1))

- P(class=1) = 5 / (5 + 5)
- P(class=0) = 5 / (5 + 5)

- **Calculate the Conditional Probabilities**
- The conditional probabilities are the probability of each input value given each class value.
- **Weather Input Variable**
- P(weather=sunny|class=go-out) = count(weather=sunny and class=go-out) / count(class=go-out)
- P(weather=rainy|class=go-out) = count(weather=rainy and class=go-out) / count(class=go-out)
- P(weather=sunny|class=stay-home) = count(weather=sunny and class=stay-home) / count(class=stay-home)
- P(weather=rainy|class=stay-home) = count(weather=rainy and class=stay-home) / count(class=stay-home)

- Plugging in the numbers we get:
- P(weather=sunny|class=go-out) = 0.8
- P(weather=rainy|class=go-out) = 0.2
- P(weather=sunny|class=stay-home) = 0.4
- P(weather=rainy|class=stay-home) = 0.6

- **Car Input Variable**
- P(car=working|class=go-out) = count(car=working and class=go-out) / count(class=go-out)
- P(car=broken|class=go-out) = count(car=brokenrainy and class=go-out) / count(class=go-out)
- P(car=working|class=stay-home) = count(car=working and class=stay-home) / count(class=stay-home)
- P(car=broken|class=stay-home) = count(car=brokenrainy and class=stay-home) / count(class=stay-home)
- Plugging in the numbers we get:
- P(car=working|class=go-out) = 0.8
- P(car=broken|class=go-out) = 0.2
- P(car=working|class=stay-home) = 0.2
- P(car=broken|class=stay-home) = 0.8.

- We can make predictions using Theorem.
- P(h|d) = (P(d|h) * P(h)) / P(d)
- Where:
- **P(h|d)** is the probability of hypothesis h given the data d. This is called the posterior probability.
- **P(d|h)** is the probability of data d given that the hypothesis h was true.
- **P(h)** is the probability of hypothesis h being true (regardless of the data). This is called the prior probability of h.
- **P(d)** is the probability of the data (regardless of the hypothesis).

- weather=sunny, car=working
- go-out = P(weather=sunny|class=go-out) * P(car=working|class=go-out) * P(class=go-out)
- go-out = 0.8 * 0.8 * 0.5
- go-out = 0.32
- We can perform the same calculation for the stay-home case:
- stay-home = P(weather=sunny|class=stay-home) * P(car=working|class=stay-home) * P(class=stay-home)
- stay-home = 0.4 * 0.2 * 0.5
- stay-home = 0.04
- We can see that 0.32 is greater than 0.04, therefore we predict “go-out” for this instance, which is correct.

Example No.	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

- We want to classify a Red Domestic SUV

P(Red|Yes), P(SUV|Yes), P(Domestic|Yes) ,

P(Red|No) , P(SUV|No), and P(Domestic|No)

```
Yes:                          No:
    Red:                          Red:
        n = 5                         n = 5
        n_c= 3                        n_c = 2
        p = .5                        p = .5
        m = 3                         m = 3
    SUV:                          SUV:
        n = 5                         n = 5
        n_c = 1                       n_c = 3
        p = .5                        p = .5
        m = 3                         m = 3
    Domestic:                     Domestic:
        n = 5                         n = 5
        n_c = 2                       n_c = 3
        p = .5                        p = .5
        m = 3                         m =3
```

$$P(Red|Yes) = \frac{3+3*.5}{5+3} = .56 \qquad P(Red|No) = \frac{2+3*.5}{5+3} = .43$$

$$P(SUV|Yes) = \frac{1+3*.5}{5+3} = .31 \qquad P(SUV|No) = \frac{3+3*.5}{5+3} = .56$$

$$P(Domestic|Yes) = \frac{2+3*.5}{5+3} = .43 \qquad P(Domestic|No) = \frac{3+3*.5}{5+3} = .56$$

```
P(Yes) * P(Red | Yes) * P(SUV | Yes) * P(Domestic|Yes)

    =  .5 * .56 * .31 * .43   = .037
```

and for $v = No$, we have

```
P(No) * P(Red | No) * P(SUV | No) * P (Domestic | No)

    = .5 * .43 * .56 * .56 = .069
```

Since 0.069 > 0.037, our example gets classified as 'NO'

- Consider following example,
- A data of 100 fruits is collected w.r.t. 3 characteristics   Long ,Sweet, Yellow

Characteristics	Banana	Mango	Any other	Total
Long	400	0	100	500
Not Long	100	300	100	500
Sweet	350	150	150	650
Not sweet	150	150	50	350
Yellow	450	300	50	800
Not Yellow	50	0	150	200
Total	500	300	200	1000

- Given a fruit characteristics as Long, sweet and yellow then predict the fruit.

- $P(B)= 500/1000= 0.5$
- $P(M)=300/1000=0.3$
- $PA)= 200/1000=0.2$
- $P(x1=Long) = 500 / 1000 = 0.50$
- $P(x2=Sweet) = 650 / 1000 = 0.65$
- $P(x3=Yellow) = 800 / 1000 = 0.80$

- P(sweet/Banana)= $\frac{P(sweet \cap Banana)}{P(Banana)} = \frac{\frac{350}{1000}}{\frac{500}{1000}} = 0.7$
- P(sweet/Mango)= $\frac{P(sweet \cap Mango)}{P(Mango)} = \frac{150/1000}{300/1000} = 0.5$
-
- P(sweet/Any other)= $\frac{P(sweet \cap Any\ other)}{P(Any\ other)} = \frac{150/1000}{200/1000} =$ 0.75
-
- P(long/Banana)= $\frac{P(long \cap Banana)}{P(Banana)} = \frac{400/1000}{500/1000} = 0.8$
- P(long/Mango)= $\frac{P(long \cap Mango)}{P(Mango)} = \frac{0/1000}{300/1000} = 0$
- P(long/Any other)= $\frac{P(long \cap Any\ other)}{P(Any\ other)} = \frac{100/1000}{200/1000} = 0.5$

- P(yellow/Banana)= $\frac{P(yellow \cap Banana)}{P(Banana)} = \frac{450/1000}{500/1000} = 0.9$
- P(yellow/Mango)= $\frac{P(yellow \cap Mango)}{P(Mango)} = \frac{300/1000}{300/1000} = 1$
- P(yellow/Any other)= $\frac{P(yellow \cap Any\ other)}{P(Any\ other)} = \frac{150/1000}{200/1000} = 0.75$

- $P(C|long, sweet, yellow) =$
$$\frac{P(c{=}Banana)\times P(Long|Banana)\times P(sweet)\times P(sweet|Banana)\times P(yellow)\times P(yellow|Banana)}{P(long)\times P(sweet)\times P(yellow)}$$
- $P(C|long, sweet, yellow) = \frac{0.5\times 0.8\times 0.7\times 0.9}{P(long)\times P(sweet)\times P(yellow)} = \frac{0.252}{0.26} = 0.96$
- $P(C|long, sweet, yellow) =$
$$\frac{P(c{=}Mango)\times P(Long|Mango)\times P(sweet)\times P(sweet|Mango)\times P(Mango)\times P(yellow|Mango)}{P(long)\times P(sweet)\times P(yellow)} = 0$$
-
-
- $P(C|long, sweet, yellow) =$
$$\frac{P(c{=}Any\ other)\times P(Long|Any\ other)\times P(sweet)\times P(sweet|Any\ other)\times P(yellow)\times P(yellow|Any\ other}{P(long)\times P(sweet)\times P(yellow)}$$
-
- $P(C|long, sweet, yellow) = \frac{0.01875}{0.26} = 0.7211$

Rec	Age	Income	Student	Credit Rating	Buy Computer
R1	<=30	High	No	Fair	No
R2	<=30	High	No	Excellent	No
R3	31…40	High	No	Fair	Yes
R4	>40	Medium	No	Fair	Yes
R5	>40	Low	Yes	Fair	Yes
R6	>40	Low	Yes	Excellent	No
R7	31…40	Low	Yes	Excellent	Yes
R8	<=30	Medium	No	Fair	No
R9	<=30	Low	Yes	Fair	Yes
R10	>40	Medium	Yes	Fair	Yes
R11	<=30	Medium	Yes	Excellent	Yes
R12	31…40	Medium	No	Excellent	Yes
R13	31…40	High	Yes	Fair	Yes
R14	>40	Medium	No	Excellent	No

- X= (age = youth, income=medium, student= yes, credit_rating=fair). Will the person buy a computer or not?

- P(C1)=P(buys_computer= yes) = 9/14= 0.643
- P(C2)= P(buys_computer= No)=5/14= 0.357
- P(age= youth| buys_computer=yes)=2/9=0.222
- P(age= youth| buys_computer=No)=3/5=0.6
- P(income = medium | buys_computer=yes)=4/9 =0.444
- P(income = medium | buys_computer=N0)=2/5=0.400
- P(student=yes| buys_computer=yes)=6/9 = 0.667
- P(student=yes| buys_computer=No)=1/5=0.2
- P(credir_rating=fair| buys_computer=Yes)=6/9 =0.667
- P(credir_rating=fair| buys_computer=No)=2/5 =0.4

- P((age = youth, income=medium, student= yes, credit_rating=fair | Buys_Computer = yes)
- = { P(age= youth| buys_computer=yes) * P(income = medium | buys_computer=yes) * P(student=yes| buys_computer=yes) *P(credir_rating=fair| buys_computer=Yes) }
- =0.22*0.44*0.667 *0.667
- =0.043
- P((age = youth, income=medium, student= yes, credit_rating=fair | Buys_Computer = No)
- = { P(age= youth| buys_computer=No) * P(income = medium | buys_computer=No) * P(student=yes| buys_computer=No) *P(credir_rating=fair| buys_computer=No) }
- =0.600*0.400*0.200*0.400=0.019

Probabilistic Models

- Two types
  - Generative
    - Naïve Bayes Classifier
  - Discriminative
    - Logistic Regression

Random Variable

- It is used to map the random processes to number
  - Flipping Coin
  - Rolling die

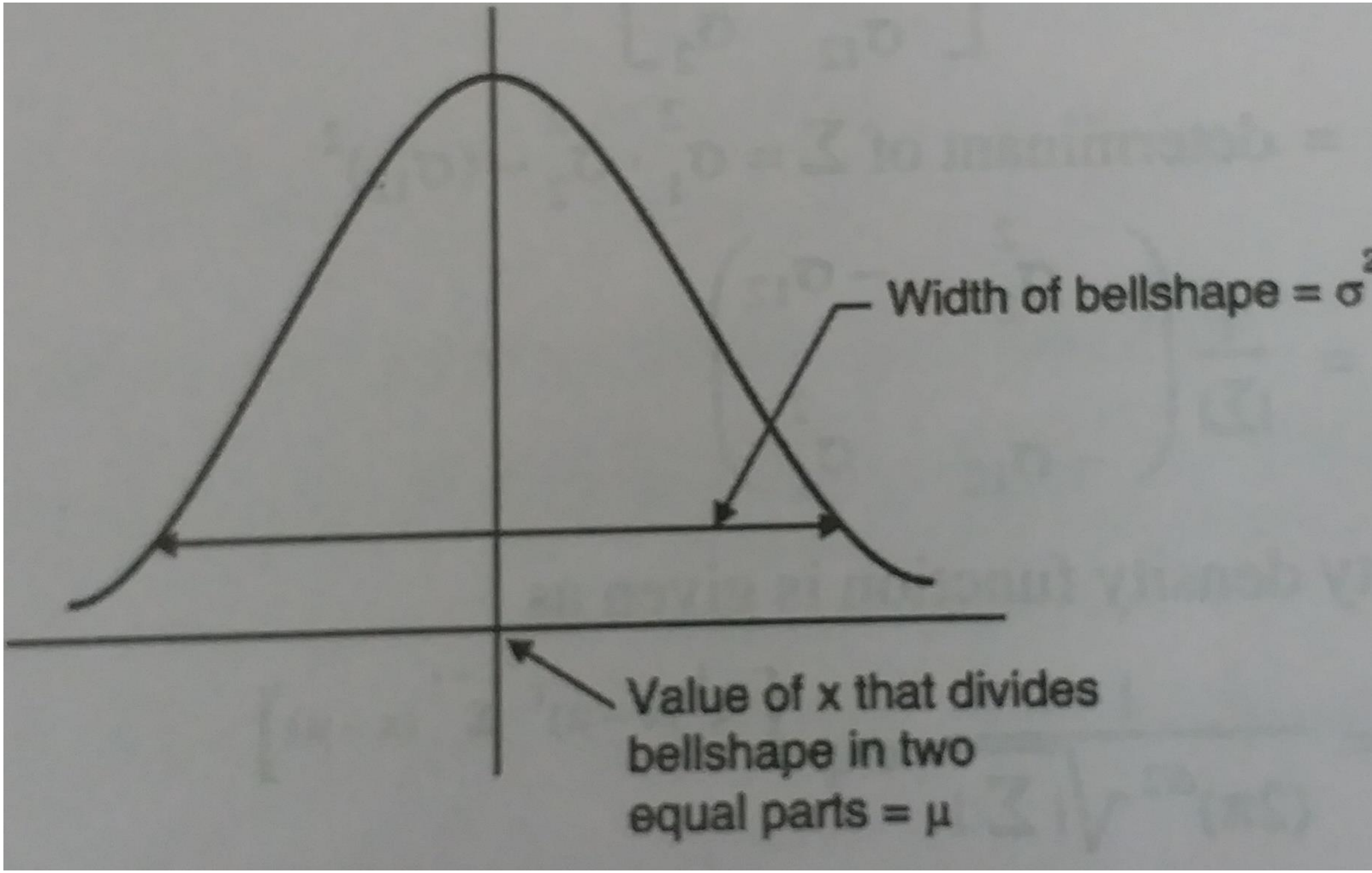
- Random Variable
  - Discrete
  - Continuous

Normal distribution

- Most commonly used in statistics and Machine Learning
- Also called as bell curves or Gaussian curves
- Normal distribution of random variable x is represented by $N(X, \mu, \sigma^2)$

- It’s probability density function (PDF) is represented as
- N(X,μ, σ^2) = $\frac{1}{\sqrt{2\pi\sigma^2}} . e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$
- X : random variable
- μ : Mean
- σ^2 : Variance of x = VAR[x]
- $\sqrt{2\pi\sigma^2}$: Normalization constant which ensures the sum of probabilities is 1

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$



Width of bellshape = σ²
Value of x that divides
bellshape in two
equal parts = μ

- A normal distribution with mean 0 and variance 1 is called standard Normal distribution

Discriminative Learning with Maximum Likelihood

- Also called as conditional models
- Discriminative function f(x,y) maps an input x to an output
- $\hat{y}$ = argmax(x,y)
- These models directly estimate posterior probabilities .

Maximum- Likelihood Estimation (MLE)

- Method of estimating the parameters of a statistical model
- Provides estimates the model's parameters.
- For X1,X2,X3…..Xn
- f_{Θ} =(X1,X2,X3…..Xn) = f(X1,X2,X3…..Xn | Θ)
- lik(Θ)= f(X1,X2,X3…..Xn | Θ)
- Maximum likelihood estimate of Θ is that value of that maximises lik(Θ)

Expectation Maximization Algorithm(EM)

- EM is an iterative method used to find maximum likelihood estimates of parameters in probabilistic models where model depend on unobserved variables
- Given a set of incomplete data, consider a set of starting parameters.
- **Expectation step (E – step):** Using the observed available data of the dataset, estimate (guess) the values of the missing data.
- **Maximization step (M – step):** Complete data generated after the expectation (E) step is used in order to update the parameters.
- Repeat step 2 and step 3 until convergence.

- **Usage of EM algorithm –**
- It can be used to fill the missing data in a sample.
- It can be used as the basis of unsupervised learning of clusters.
- It can be used for the purpose of estimating the parameters of Hidden Markov Model (HMM).
- It can be used for discovering the values of latent variables.

- **Advantages of EM algorithm –**
- It is always guaranteed that likelihood will increase with each iteration.
- The E-step and M-step are often pretty easy for many problems in terms of implementation.
- Solutions to the M-steps often exist in the closed form.

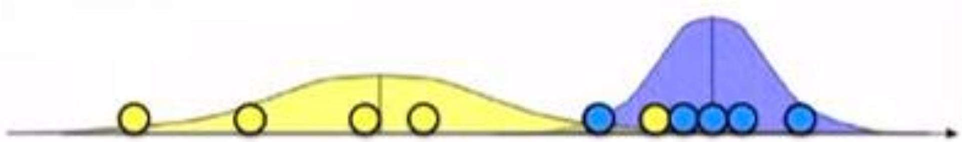
- **Disadvantages of EM algorithm –**
- It has slow convergence.
- It makes convergence to the local optima only.
- It requires both the probabilities, forward and backward (numerical optimization requires only forward probability).

- Let X1,X2,X3,…. , Xn are the observations

- K= 2 with Gaussians with unknown μ ,σ^2
- Estimation is trivial if we know the source

$$\mu_b = \frac{x_1 + x_2 + \ldots + x_{n_b}}{n_b}$$

$$\sigma_b^2 = \frac{(x_1 - \mu_1)^2 + \ldots + (x_n - \mu_n)^2}{n_b}$$



- What if we don’t know the source????

$$P(b \mid x_i) = \frac{P(x_i \mid b)P(b)}{P(x_i \mid b)P(b) + P(x_i \mid a)P(a)}$$

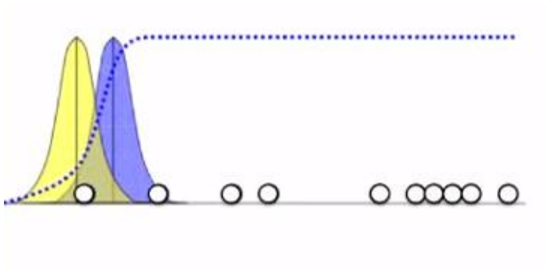
$$P(x_i \mid b) = \frac{1}{\sqrt{2\pi\sigma_b^2}} \exp\left(-\frac{(x_i - \mu_b)^2}{2\sigma_b^2}\right)$$

- We can guess if we know μ ,σ^2

Expected Maximization (EM)Algorithm

- We need μ_a ,$\sigma^2{}_a$ and μ_b ,$\sigma^2{}_b$ to guess the source
- And we need the source to estimate μ_a ,$\sigma^2{}_a$ and μ_b ,$\sigma^2{}_b$
- EM algorithm is used to estimate the parameters
  - Start with 2 randomly placed Gaussians (μ_a ,$\sigma^2{}_a$) and(μ_b ,$\sigma^2{}_b$)
  - For each point estimate $P(b|x_i)$ to guess the source
  - Adjust μ_a ,$\sigma^2{}_a$ and μ_b ,$\sigma^2{}_b$ to fit the points assigned to them
  - Iterate until it converge

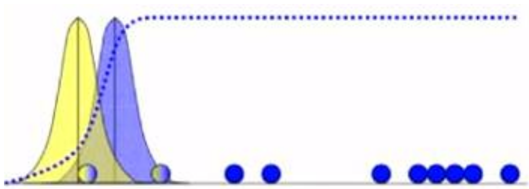
Example



$$P(x_i \mid b) = \frac{1}{\sqrt{2\pi\sigma_b^2}} \exp\left(-\frac{(x_i - \mu_b)^2}{2\sigma_b^2}\right)$$

$$b_i = P(b \mid x_i) = \frac{P(x_i \mid b)P(b)}{P(x_i \mid b)P(b) + P(x_i \mid a)P(a)}$$

$$a_i = P(a \mid x_i) = 1 - b_i$$



$$\mu_b = \frac{b_1 x_1 + b_2 x_2 + \ldots + b_n x_{n_b}}{b_1 + b_2 + \ldots + b_n}$$

$$\sigma_b^2 = \frac{b_1(x_1 - \mu_1)^2 + \ldots + b_n(x_n - \mu_n)^2}{b_1 + b_2 + \ldots + b_n}$$