Bayes for Beginners

Presenters: Shuman ji & Nick Todd

Statistic Formulations.

$P(A)$: probability of event A occurring

$P(A|B)$: probability of A occurring given B occurred

$P(B|A)$: probability of B occurring given A occurred

$P(A,B)$: probability of **A and B** occurring simultaneously (joint probability of A and B)

Joint probability of A and B

$$P(A,B) = P(A|B)*P(B) = P(B|A)*P(A)$$

Bayes Rule

- True Bayesians actually consider conditional probabilities as more basic than joint probabilities . It is easy to define P(A|B) without reference to the joint probability P(A,B). To see this note that we can rearrange the conditional probability formula to get:

- P(A|B) P(B) = P(A,B)

by symmetry:

- P(B|A) P(A) = P(A,B)

- It follows that:

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

- which is the so-called ***Bayes Rule.***
- Thus, Bayes Rule is a simple mathematical formula used for calculating conditional probabilities

Bayesian Reasoning

ASSUMPTIONS

$P(A)$ =1% of women aged forty who participate in a routine screening have breast cancer

$P(B|A)$=80% of women with breast cancer will get positive tests

9.6% of women without breast cancer will also get positive tests

EVIDENCE

A woman in this age group had a positive test in a routine screening

PROBLEM

What's the probability that she has breast cancer?

= proportion of cancer patients with positive results, within the group of All patients with positive results

---→ $P(A|B)$

$P(B)$= proportion of all patients with positive results

Bayesian Reasoning

ASSUMPTIONS

- 100 out of 10,000 women aged forty who participate in a routine screening have breast cancer
- 80 of every 100 women with breast cancer will get positive tests
- 950 out of 9,900 women without breast cancer will also get positive tests

PROBLEM

- If 10,000 women in this age group undergo a routine screening, about what fraction of women with positive tests will actually have breast cancer?

Bayesian Reasoning

Before the screening:

100 out of 10000 women with breast cancer

9,900 out of 10000 women without breast cancer

After the screening:

A = 80 out of 10000 women with breast cancer and positive test

B = 20 out of 10000 women with breast cancer and negative test

C = 950 out of 10000 women without breast cancer and positive test

D = 8,950 out of 10000 women without breast cancer and negative test

All patients that has positive test result = A+C

Proportion of cancer patients with positive results, within the group of ALL patients with positive results:

A/(A+C) = 80/(80+950) = 80/1030 = 0.078 = 7.8%

Bayesian Reasoning

Prior Probabilities:

100/10,000 = 1/100 = 1% = p(A)

9,900/10,000 = 99/100 = 99% = p(~A)

Conditional Probabilities:

A = 80/10,000 = (80/100)*(1/100) = p(B|A)*p(A) = 0.008

B = 20/10,000 = (20/100)*(1/100) = p(~B|A)*p(A) = 0.002

C = 950/10,000 = (9.6/100)*(99/100) = p(B|~A)*p(~A) = 0.095

D = 8,950/10,000 = (90.4/100)*(99/100) = p(~B|~A) *p(~A) = 0.895

Rate of cancer patients with positive results, within the group of ALL patients with positive results:

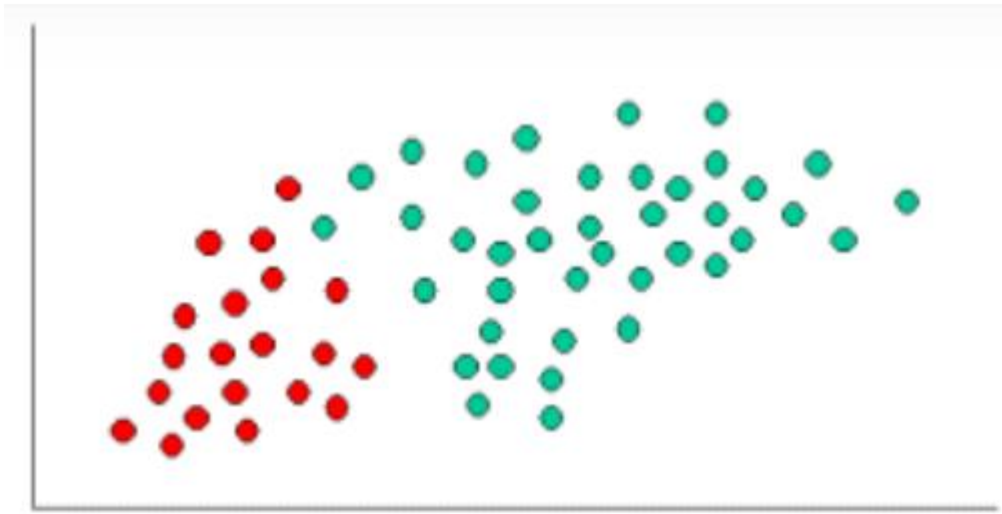
P(A|B) = P(B|A) * P(A)/ P(B) = 0.008/(0.008+0.095) = 0.008/0.103 = 0.078 = 7.8%

Another example

- Suppose that we are interested in diagnosing cancer in patients who visit a chest clinic:
- Let A represent the event "Person has cancer"
- Let B represent the event "Person is a smoker"
- We know the probability of the prior event $P(A)=0.1$ on the basis of past data (10% of patients entering the clinic turn out to have cancer). We want to compute the probability of the posterior event $P(A/B)$. It is difficult to find this out directly. However, we are likely to know $P(B)$ by considering the percentage of patients who smoke – suppose $P(B)=0.5$. We are also likely to know $P(B/A)$ by checking from our record the proportion of smokers among those diagnosed. Suppose $P(B/A)=0.8$.
- We can now use Bayes' rule to compute:
- $P(A/B) = (0.8 * 0.1)/0.5 = 0.16$
- Thus we found the proportion of cancer patients who are smokers.

Bayes Classifier

Naïve Bayes Classifier



Given the following data, labelled as Red and Green. Predict the class of a new point

Bayes Classifier

Priori Probability

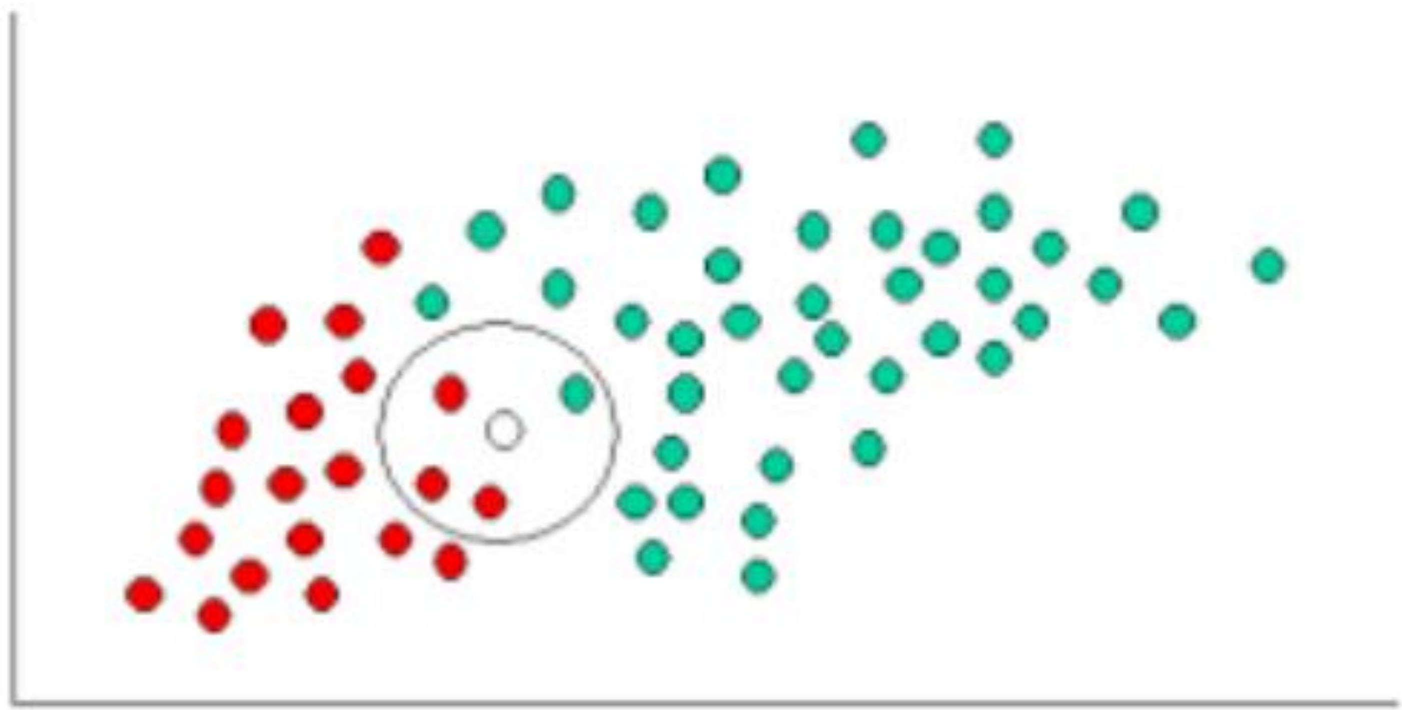
Prior Probability of `GREEN` : `number of GREEN objects / total number of objects`

Prior Probability of `RED` : `number of RED objects / total number of objects`

Prior Probability for `GREEN` : `40 / 60`

Prior Probability for `RED` : `20 / 60`

Bayes Classifier



The likely hood of a point is defined by its neighbours

Bayes Classifier

$$\text{Likelihood of } X \text{ given GREEN} \propto \frac{\text{Number of GREEN in the vicinity of } X}{\text{Total number of GREEN cases}}$$

$$\text{Likelihood of } X \text{ given RED} \propto \frac{\text{Number of RED in the vicinity of } X}{\text{Total number of RED cases}}$$

$$\text{Probability of } X \text{ given GREEN} \propto \frac{1}{40}$$

$$\text{Probability of } X \text{ given RED} \propto \frac{3}{20}$$

Bayes Classifier

$$\text{Posterior probability of } X \text{ being GREEN} \propto$$

$$\text{Prior probability of GREEN} \times \text{Likelihood of } X \text{ given GREEN}$$

$$= \frac{4}{6} \times \frac{1}{40} = \frac{1}{60}$$

$$\text{Posterior probability of } X \text{ being RED} \propto$$

$$\text{Prior probability of RED} \times \text{Likelihood of } X \text{ given RED}$$

$$= \frac{2}{6} \times \frac{3}{20} = \frac{1}{20}$$

Bayes Classifier

Second Example

Let's say that we have data on 1000 pieces of fruit. They happen to be **Banana**, **Orange** or some **Other Fruit**. We know 3 characteristics about each fruit:

1. Whether it is Long
2. Whether it is Sweet and
3. If its color is Yellow.

Type	Long	Not Long	Sweet	Not Sweet	Yellow	Not Yellow	Total
Banana	400	100	350	150	450	50	500
Orange	0	300	150	150	300	0	300
Other Fruit	100	100	150	50	50	150	200
Total	500	500	650	350	800	200	1000

Let's say that we are given the properties of an unknown fruit, and asked to classify it. We are told that the fruit is Long, Sweet and Yellow. Is it a Banana? Is it an Orange? Or Is it some Other Fruit?

Bayes Classifier

Second Example

The so-called "Prior" probabilities.

```
P(Banana)       = 0.5 (500/1000)
P(Orange)       = 0.3
P(Other Fruit) = 0.2
```

Probability of "Evidence"

```
p(Long)   = 0.5
P(Sweet)  = 0.65
P(Yellow) = 0.8
```

Probability of "Likelihood"

```
P(Long|Banana) = 0.8
P(Long|Orange) = 0  [Oranges are never long in all the fruit we have seen.]
 ....

P(Yellow|Other Fruit)     =  50/200 = 0.25
P(Not Yellow|Other Fruit) = 0.75
```

Bayes Classifier

Second Example

```
P(Banana|Long, Sweet and Yellow)
      P(Long|Banana) * P(Sweet|Banana) * P(Yellow|Banana) * P(banana)
    = _______________________________________________________________
                      P(Long) * P(Sweet) * P(Yellow)

    = 0.8 * 0.7 * 0.9 * 0.5 / P(evidence)

    = 0.252 / P(evidence)
```

Bayes Classifier

Second Example

```
P(Orange|Long, Sweet and Yellow) = 0


P(Other Fruit|Long, Sweet and Yellow)
      P(Long|Other fruit) * P(Sweet|Other fruit) * P(Yellow|Other fruit) * P(Other Fruit)
    = _____________________________________________________________________________________
                                           P(evidence)


    = (100/200 * 150/200 * 50/200 * 200/1000) / P(evidence)

    = 0.01875 / P(evidence)
```