Naïve Bayes

- Sonal Ghanshani

Naïve Bayes

- Applied to multi class classification problem
- This is a probabilistic model which assumes conditional independence between features.
- Given a set of features, Naive Bayes classifier is used to predict a class using probability.

Probability

Probability

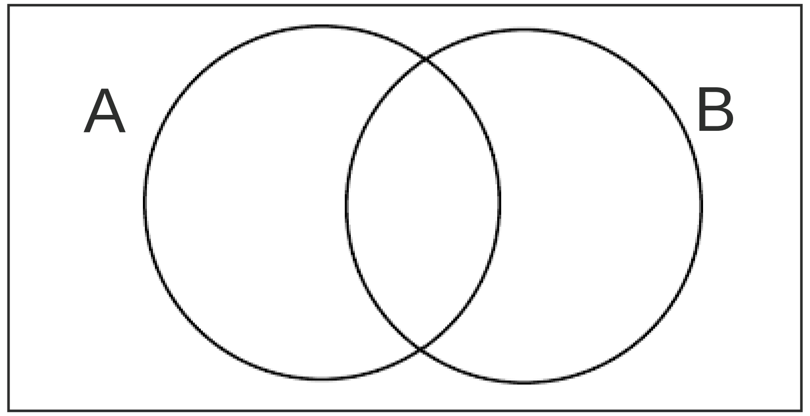
- Probability is the likelihood of an event occurring.
- The probability of event A or event B is denoted by the event $A \cup B$. Its probability is given by $P(A \cup B)$
- The probability of event A and event B is denoted by the event $A \cap B$. Its probability is given by $P(A \cap B)$. For disjoint events, A and B, $P(A \cap B)$=0
- If A is an event then A not occurring is also an event which is denoted by A'

Disjoint

- Two events are disjoint if they cannot occur at the same time
  - During the toss of a coin, getting head and tail are two disjoint events.
  - While drawing a card from the deck of 52, obtaining an Ace and a King are also disjoint events.
  - When a die is rolled, the probability of getting 3 and a multiple of 2 are two mutually exclusive events.

Rules of Probability

- For two disjoint events, $P(A \cup B) = P(A) + P(B)$
- Rule of Subtraction
  - Probability of an event not occurring is given by $P(A') = 1 - P(A)$
  - Directly derived from $P(A) + P(A') = 1$
- Rule of Addition
  - Probability of Event $A \cup B$ is given by $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $P(A' \cup B') = 1 - P(A \cup B)$



Conditional Probability

- The probability that Event A will occur given that Event B has already occurred. Denoted by $P(A|B)$
  - P(A|B) = the probability of A being true, given that we know that B is true
    - H = “I have a headache”
    - F = “Coming down with flu”
    - P(H) = 1/10
    - P(F) = 1/40
    - P(H/F) = 1/2
  - Headaches are rare and flu even rarer, but if you got flu, there is a 50-50 chance you’ll have a headache

Rules of Probability

- Rule of Multiplication
  - $P(A \cap B) \ = \ P(A).P(B|A) \ = \ P(B).P(A|B)$
  - $P(A \cap B) \ = \ P(A).P(B)$ if the events are independent of each other
- Conditional Probability,
  - $P(A|B) = \frac{P(A \cap B)}{P(B)} \ = \frac{P(A).P(B \mid A)}{P(B)}$

Bayes Theorem

- Bayes' rule is a formula that extends the use of the law of conditional probabilities to allow revision of original probabilities with new information
- Derived from the conditional probabilities is the Bayes Theorem

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

Bayes Theorem

- The most common form of colour-blindness is a sex-linked hereditary condition caused by a defect on the X chromosome. Thus, it is more common in males than females; 7% of males are colour-blind but only 0.5% of females are colour-blind. In a certain population, 50% are male and 50% are female. Find the percentage of colour-blind persons that are ma

M = Male   C - Colour Blind   F = Female

$P(C|M) = 0.07$ $P(M) = 0.5$

$P(E|F) = 0.005$ $P(F) = 0.5$

$P(M|C) = ?$

Using Bays Theorem,

$$P(M|C) = \frac{P(C|M) \cdot P(M)}{P(C)}$$

$$= \frac{P(C|M) \cdot P(M)}{P(C|M) \cdot P(M) + P(C|F) * P(F)}$$

$$= \frac{0.07 \times 0.5}{(0.07 \times 0.5) + (0.005 \times 0.5)}$$

$$= \frac{0.035}{0.035 + 0.0025}$$

$$= \frac{0.035}{0.0375}$$

$$= 0.9333$$

Bayes Theorem

- You have to decide whether or not to study hard for your Stats final. The professor tells you that, in the past, 75% of the A's belong to people who study hard, and 20% of the non-As belong to people who study hard. Furthermore, experience shows that about 40% of people get A's on the final.
  - What is the probability of getting an A, given that you study hard?

A: Getting A

$\bar{A} \equiv$ Not getting A

$SH \equiv$ studying hard

$\overline{SH} \equiv$ Not Studying hard

$P(A) = 40\% = 0.4$; $P(\bar{A}) = 0.6$

$P(SH|A) = 0.75$ $\qquad$ $P(SH|\bar{A}) = 0.2$

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

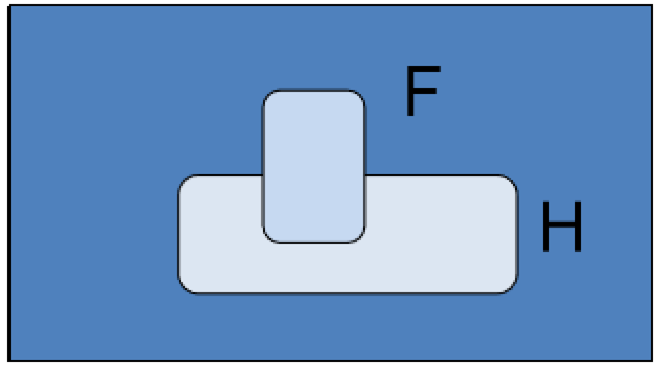
$$P(SH) = P(SH|A)\cdot P(A) + P(SH|\bar{A})\cdot P(\bar{A})$$
$$= (0.75)(0.4) + (0.2 \times 0.6)$$
$$= 0.3 + 0.12$$
$$= 0.42$$

$$P(A|SH) = \frac{0.75 \times 0.4}{0.42} = 0.7142$$

Naïve Bayes Classifier

Naïve Bayes Classifier

- P(A|B) = the probability of A being true, given that we know that B is true



H = "I have a headache"
F = "Coming down with flu"

P(H) = 1/10
P(F) = 1/40
P(H/F) = 1/2

Headaches are rare and flu even rarer, but if you got flu, there is a 50-50 chance you'll have a headache.

Naïve Bayes Classifier

- Consider,
  - P(Y): Prob (Loan Sanctioned)
  - P(X): Prob (Customer is educated)
  - P(X|Y): Prob(Educated | Loan Sanctioned)

  - We are interested in P(Y|X): Prob(Educated | Loan Sanctioned)

Educated	Loan Sanctioned
1	1
0	0
1	1
1	1
0	0
0	0
0	1
1	1
0	0
0	0
0	1
1	1
0	0
0	1
0	0
0	0
1	1
0	0
1	1
1	1

Another Example of the Naïve Bayes Classifier

Day	Outlook	Temperature	Humidity	Wind	Play Tennis
Day1	Sunny	Hot	High	Weak	No
Day2	Sunny	Hot	High	Strong	No
Day3	Overcast	Hot	High	Weak	Yes
Day4	Rain	Mild	High	Weak	Yes
Day5	Rain	Cool	Normal	Weak	Yes
Day6	Rain	Cool	Normal	Strong	No
Day7	Overcast	Cool	Normal	Strong	Yes
Day8	Sunny	Mild	High	Weak	No
Day9	Sunny	Cool	Normal	Weak	Yes
Day10	Rain	Mild	Normal	Weak	Yes
Day11	Sunny	Mild	Normal	Strong	Yes
Day12	Overcast	Mild	High	Strong	Yes
Day13	Overcast	Hot	Normal	Weak	Yes
Day14	Rain	Mild	High	Strong	No

Another Example of the Naïve Bayes Classifier

The weather data, with counts and probabilities

outlook			temperature			humidity			windy			play	
	yes	no		yes	no		yes	no		yes	no	yes	no
sunny	2	3	hot	2	2	high	3	4	false	6	2	9	5
overcast	4	0	mild	4	2	normal	6	1	true	3	3		
rainy	3	2	cool	3	1								
sunny	2/9	3/5	hot	2/9	2/5	high	3/9	4/5	false	6/9	2/5	9/14	5/14
overcast	4/9	0/5	mild	4/9	2/5	normal	6/9	1/5	true	3/9	3/5		
rainy	3/9	2/5	cool	3/9	1/5								

A new day

outlook	temperature	humidity	windy	play
sunny	cool	high	true	?

Another Example of the Naïve Bayes Classifier

- Likelihood of yes

$$= \frac{2}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{9}{14} = 0.0053$$

- Likelihood of no

$$= \frac{3}{5} \times \frac{1}{5} \times \frac{4}{5} \times \frac{3}{5} \times \frac{5}{14} = 0.0206$$

- Therefore, the prediction is No

Model Validation

Accuracy Matrix

		Predicted class	
		P	*N*
Actual Class	*P*	True Positives (TP)	False Negatives (FN)
	N	False Positives (FP)	True Negatives (TN)

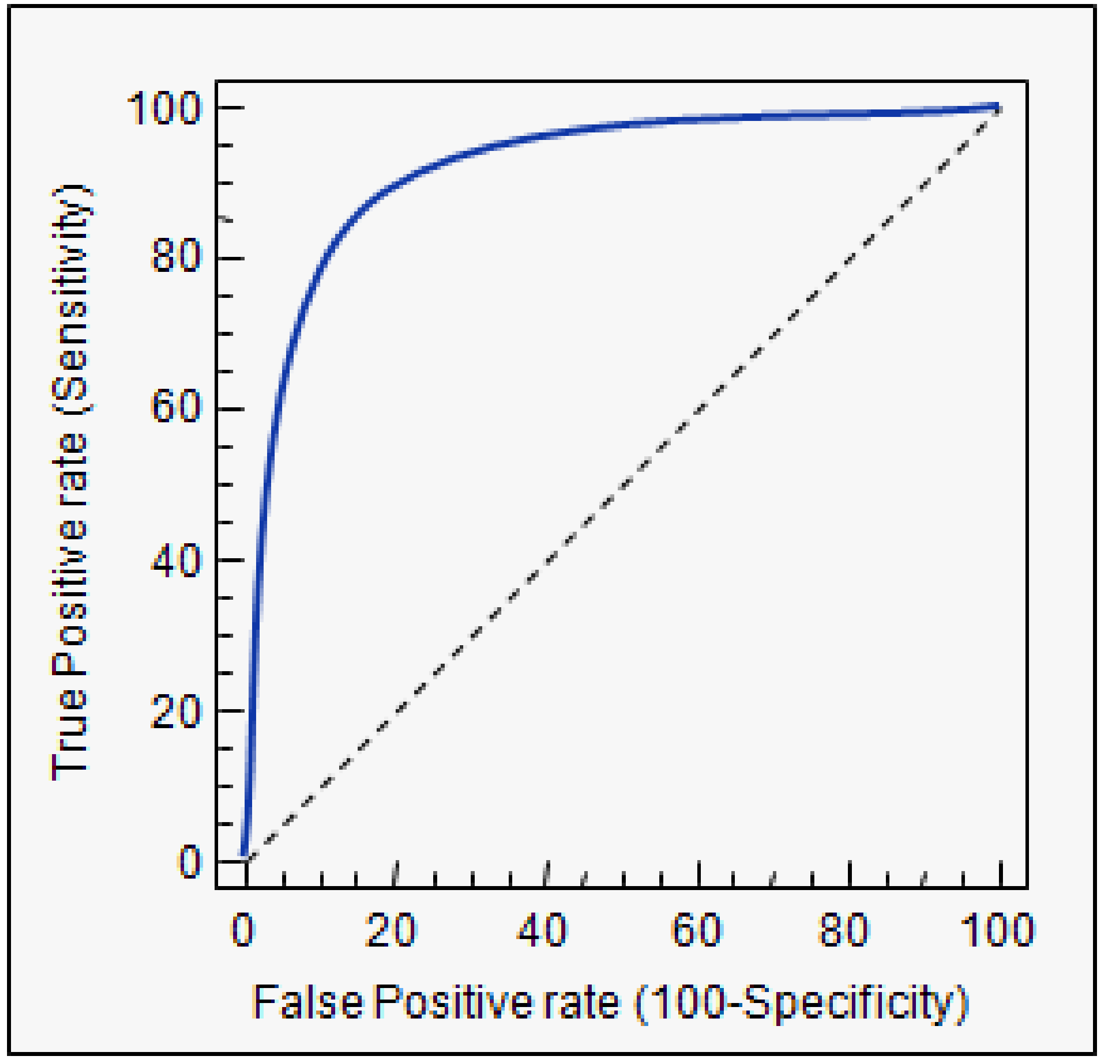
Sensitivity, recall, hit rate, or true positive rate (TPR)

$$\mathrm{TPR} = \frac{\mathrm{TP}}{P} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

Specificity or true negative rate (TNR)

$$\mathrm{TNR} = \frac{\mathrm{TN}}{N} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}$$

ROC - AUC


100
80
60
40
20
0
True Positive rate (Sensitivity)
0
20
40
60
80
100
False Positive rate (100-Specificity)

Application

Things We’d Like To Do

- Spam Classification
  - Given an email, predict whether it is spam or not
- Medical Diagnosis
  - Given a list of symptoms, predict whether a patient has disease X or not
- Weather
  - Based on temperature, humidity, etc… predict if it will rain tomorrow
- Digit Recognition