Machine Learning

Naïve Bayes Classifier

Naïve Bayes Classifier

- Naïve Bayes Classifier is a probabilistic model
- It is based on Bayes theorem. It is called “naïve” due to the assumption that the features in the dataset are independent which may not be true in the real world.
- Despite this, the classifier performs well.
- It factors all available evidence in form of predictors into the Naïve Bayes rule to obtain more accurate probability
- Results in terms of class membership probabilities, such as the probability that a given sample belongs to a particular class.

Naïve Bayes Classifier

- Bayes theorem is based on the concept of conditional probability.

- Conditional probability of an event, is the probability that the event will occur, given that something else has already occurred. For example, probability that an email is likely to be a spam given that the email contains words “lottery”

Probability Basics

- Probability is chance of occurrence of an event. It's values is always between 0 and 1
- P(A) = p / q $P(H) = \frac{1}{2}$ H T
- Where
- p = number of ways that are favorable to the occurrence of A
- q = the total number of outcomes
- Questions:
  - What is the probability of getting a Head in toss of a coin?
  - What is the probability of getting score 2 when a die is rolled
  - What is the probability of getting total score of 3 when 2 dice are rolled

Probability Basics

Joint Probability – is the probability of multiple events occurring together. For example

- Probability of drawing an ace from a deck of cards
  4/52

- Probability of drawing a red colour card from a deck of cards
  26/52

- Probability of drawing a red colour ace
  2 / 52

P(Red and Ace)

$P(Red \cap Ace) = P(Red) * P(Ace)$

$= \frac{\cancel{4}\,2}{52} \times \frac{\cancel{26}}{\cancel{52}\,2} = \frac{2}{52}$

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

Probability Basics

- Conditional Probability:
  - If you play on 20 of 30 day, then what is probability that you will play on a day?
  - Ans: P(play) = 20/30 = 2/3

  - If it is raining on the day, is the probability that you will play be the same?

  - Given that it is raining (an event has occurred), what is the probability that you will play? This is where conditional probability.

  - This is denoted as **P(play | raining)**

P(play given that it is raining)

Probability Basics

Conditional Probability – it is the probability that an event given another event has occurred. For example,

- **Given the card drawn is red** (an event has occurred)
- **What is the probability it is an ace** (event not yet observed)?
- Since the card is red, there are 26 red cards. Of these 26 possible values we are interested in aces which is 2. Thus the conditional probability that the card is a ace given red card is 2 /26 $\rightarrow$ $P(\text{Ace} \mid \text{Card is Red}) = P(\text{Red} \cap \text{Ace}) / P(\text{Red})$

  $= \frac{2}{52} / \frac{1}{2} = \frac{2}{26}$
- Compare this with joint probability of red ace (2/52).
- Given an event has occurred, the probability of the other event can be revised using following

  P(A | B) = P(A ∩ B) / P(B)

$P(A \cap B) = P(A \mid B) * P(B) \leftarrow$ A dep. on B

$P(A \cap B) = P(A) * P(B) \leftarrow$ A & B independent

Conditional Probability

contingency table

	India Win	India Lose	Total
	60	40	100

1.
- What is the probability of India Win (IW)?
  - P(IW) = 60/100

2.
- What is the probability of Virat scoring a century (VC)?
  - P(VC) = 14/100

3.
- What is the probability of India win and Virat scoring a century?
  - P(IW and VC) = P(IW ∩ VC) = 10/100

4.
- What is the probability of India Win, given that Virat has score century?
  - P(IW | VC) = 10/14
- P(IW | VC) = 10/14 = (10/100) / (14/100) = P(IW ∩ VC) / P (VC)
- Thus, P(A | B) = P(A ∩ B) / P(B)

$$P(IW \mid VC) = \frac{P(VC \mid IW) * -}{10/16}$$

VC	IW
Y	Y
Y	Y
N	Y
N	N
Y	N
Y	Y
:	:

Bayes theorem

$$P(\text{Hit Mineral} \mid \text{Test Pos}) = P(\text{Test Pos} \mid \text{Hit Min}) \times$$

- P(A | B) = P(A ∩ B) / P(B)
- P(A ∩ B) = P(A | B) * P(B)
- But P(A ∩ B) = P(B ∩ A)
- Therefore,
- P(A | B) * P(B) = P(B | A) * P(A)
- $\boxed{\textbf{P(A | B) = P(B | A) * P(A) / P(B)}}$

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

Training Data

... X_n #Preg	Y Diab
1	Y
0	N
0	N
2	N
1	Y
4	Y
5	Y

$$\text{posterior probability} = \frac{\text{conditional probability} \cdot \text{prior probability}}{\text{evidence}}$$

$$P(\text{Diab} = Y \mid \#\text{Preg} = 5) = P(\text{Preg} = 5 \mid \text{Diab} = Y) \times \ldots =$$

Naïve Bayes Classifier

$$\text{posterior probability} = \frac{\text{conditional probability} \cdot \text{prior probability}}{\text{evidence}}$$

- **P(A | B) = P(B | A) * P(A) / P(B)**

- In the context of a classification problem, the Bayes theorem equation is: What is the probability that a particular object belongs to a class given its observed feature values?

ω_1 = Cancer
ω_2 = No Cancer

$P(\text{Cancer} \mid x_i)$

$$\boxed{P(\omega_j \mid \mathbf{x}_i) = P(\mathbf{x}_i \mid \omega_j) \cdot P(\omega_j)} \Big/ \cancel{P(\mathbf{x}_i)}$$

$P(\text{No Cancer} \mid x_i)$

- $\mathbf{x}_i$ be the feature vector of sample i, $i \in \{1, 2, ..., n\}$,
- ω_j be the notation of class j, $j \in \{1, 2, ..., m\}$,
- and $P(\mathbf{x}_i \mid \omega_j)$ be the probability of observing sample $\mathbf{x}_i$ given that is belongs to class ω_j.

Naïve Bayes Classifier

- Class is predicted based on maximum posterior probability of class
- If P(default | x_i) > P(not-default | x_i), then classify as “default” else classify as “not-default”.
- If P(cancer | x_i) > P(not-cancer | x_i), then classify as “cancer” else classify as “healthy”.

Naïve Bayes Classifier

- For a d-dimensional features (x_1, x_2… x_d), the class conditional probability can be calculated as follows

x_1 x_2 .... $P(x_1 \,\&\, x_2 \,\&\, x_3)$

x_1 x_2 .. x_n

2 - -

$$P(\omega_j \mid \mathbf{x}_i) = \frac{P(\mathbf{x}_i \mid \omega_j) \cdot P(\omega_j)}{\cancel{P(\mathbf{x}_i)}}$$

$$P(\mathbf{x}_i \mid \omega_j) = P(x_1 \mid \omega_j) \cdot P(x_2 \mid \omega_j) \cdot \ldots \cdot P(x_d \mid \omega_j) = \prod_{k=1}^{d} P(x_k \mid \omega_j)$$

- Bayes classifier formula assumes independence of features.

$P(\text{Diab} = Y \mid \text{Prg} = 5 \,\&\, BP = - \,\&\, - \,.) = P(\text{Prg} = 5 \,\&\, x_2 \,\&\, .. \mid D)$

$= P(\text{Prg} = 5 \mid D) \cdot P(BP = \mid D)$

Exercise

Naïve Bayes Classifier

X_s

Gaussian NB	Continuous
Multinomial NB	Discrete
Bernolli NB	Binary (0 or 1)

Missing Data →

1. Delete (Not the first choice)
2. Replace with Median, Mean, Mode
3. Imputation → KNN - R - MICE

x_n Regression to predict missing X.

?
?
?

4. Special logic

x_1 x_2 . .

20 - ? 15 - ?.

10^{-} 9 & 9 ..