NAÏVE BAYES CLASSIFIER

Class conditional Independence.

NAÏVE BAYES CLASSIFIER.

Also known as **Idiot's Bayes** or **simple bayesian classifier/ statistical classifier.**

Makes use of the **Bayes theorem** to compute probabilities of class membership, given specific evidence.

BAYES THEOREM

- At the heart of this approach is the Bayes theorem:

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)}$$

- Theorem calculates the probability of a hypothesis (H) given some evidence (E), or **posterior probability P(H|E)**
- For example, it can **calculate the probability that someone would develop diabetes** *given evidence of a family history of diabetes.*
- The **hypothesis corresponds to the response variable** in the other methods.
- The theorem makes use of this **posterior probability** of the evidence given the hypothesis, or **P(E|H).**

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)}$$

BAYES THEOREM

- Using the same example, **the probability of someone having a family history of diabetes** can also be calculated given the *evidence that the person has diabetes* and would be an example of P(E|H).
- The formula also makes use of **two prior probabilities:**
  - **The probability of the hypothesis P(H)**, and
  - **The probability of the evidence P(E).**
- These ***probabilities are not predicated*** on the presence of any evidence

P (Head)

P (Tail)

PROBABILITY | ½ = 50 %

4 Queens
52 cards

P (Queen)

PROBABILITY

1/13

Queen of Diamond

52 cards
13 Diamonds
1 Queen

PROBABILITY

$P(Q \mid D) = 1/13$

Occurred

INDEPENDENCE ASSUMPTION.

- In strict use of the Bayes theorem for **multiple independent variables** each having multiple possible values becomes challenging in practical situations.
- Using this formula directly would result in a large number of computations.
- Also, the **training data would have to cover all of these situations**, which also makes its application **impractical.**
- The naive Bayes approach **uses a simplification** which results in a **computationally feasible series of calculations.**
- The method **assumes that the independent variables are independent despite the fact** that **this is rarely** the case.
- Even with this **overly optimistic assumption**, the method is useful as a **classification modelling** method in many situations

INDEPENDENCE ASSUMPTION.

- Naive Bayesian classifiers assumes that:
  - Effect of an attribute value on a given class is independent of the values of the other attributes.
  - This assumption is called **class conditional independence**

CLASSIFICATION PROCESS.

- **Observation (X):**

  BP = high;

  Weight = above;

  FH = yes;

  Age = 50 +

- **Objective:** To classify this individual as prone to developing or not prone to ***developing diabetes*** given the factors described.

Training data

TABLE 4.19 Diabetes Data Set to Illustrate the Naive Bayes Classification

Blood pressure	Weight	Family history	Age	Diabetes
Average	Above average	Yes	50+	1
Low	Average	Yes	0 50	0
High	Above average	No	50+	1
Average	Above average	Yes	50+	1
High	Above average	Yes	50+	0
Average	Above average	Yes	0 50	1
Low	Below average	Yes	0 50	0
High	Above average	No	0 50	0
Low	Below average	No	0 50	0
Average	Above average	Yes	0 50	0
High	Average	No	50+	0
Average	Average	Yes	50+	1
High	Above average	No	50+	1
Average	Average	No	0 50	0
Low	Average	No	50+	0
Average	Above average	Yes	0 50	1
High	Average	Yes	50+	1
Average	Above average	No	0 50	0
High	Above average	No	50+	1
High	Average	No	0 50	0

CLASSIFICATION PROCESS.

- Calculate **P(diabetes=1|X)** and the **P(diabetes=0|X)** is the next step
- The individual will be assigned to the class, either **has (diabetes=1)** or **has not (diabetes=0)**, based on the **highest probability value.**

$$P(diabetes = 1|X) = \frac{P(X|diabetes = 1)P(diabetes = 1)}{P(X)} \qquad \textbf{(1)}$$

$$P(diabetes = 0|X) = \frac{P(X|diabetes = 0)P(diabetes = 0)}{P(X)} \qquad \textbf{(2)}$$

- Since **P(X) is the same in both equations**, only

  P(X| diabetes=1)P(diabetes=1) and **P(X| diabetes=0)P(diabetes=0)**
- To calculate **P(diabetes=1)**

$$= \frac{\text{Number of observations with diabetes=1}}{\text{Total number of observations}}$$

= 9/20

= **0.45** **(3)**

CLASSIFICATION PROCESS.

BP = high;
Weight = above;
FH = yes;
Age = 50 +

To calculate **P(diabetes=0)**

$$= \frac{\text{Number of observations with diabetes=0}}{\text{Total number of observations}}$$

$= 11/20$

$=$ **0.55** **(4)**

Since this approach assumes that the **independent variables are independent**,

P(X |diabetes=1) = Product of conditional probability for each value of X:

P(X| diabetes = 1) = P(BP = high | diabetes = 1)
x P(weight = above | diabetes = 1)
x P(FH = yes | diabetes = 1)
x P(age = 50 | diabetes = 1)

CLASSIFICATION PROCESS.

- P(BP=high | diabetes=1) = $\frac{\text{No. of observations with BP high and diabetes=1}}{\text{No. of observations where diabetes=1}}$

  P(BP = high | diabetes = 1) = 4/9 = 0:44
  P(weight = above | diabetes = 1) = 7/9 = 0:78
  P(FH = yes | diabetes = 1) = 6/9 = 0:67
  P(age = 50+ | diabetes = 1) = 7/9 = 0:78

- Using these probabilities, the probability of X given diabetes=1 is calculated:

  P(X | diabetes = 1) = P(BP = high | diabetes = 1)
  x P(weight = above | diabetes = 1)
  x P(FH = yes | diabetes = 1)
  x P(age = 50 | diabetes = 1)
  = 0.44 x 0.78 x 0.67 x 0.78
  = 0.179

CLASSIFICATION PROCESS.

- Using the values for P(X| diabetes=1) and P(diabetes =1), the product

  P(X | diabetes=1)P(diabetes 1)

  = 0.179 x 0. 45

  = 0.081

- Similarly, value for **P(X | diabetes=0)P(diabetes=0)** can be calculated:

  P(X | diabetes = 0) = P(BP = high | diabetes = 0)

  x P(weight = above | diabetes = 0)

  x P(FH = yes | diabetes = 0)

  x P(age = 50+ | diabetes = 0)

CLASSIFICATION PROCESS.

- Using the following probabilities, based on counts from Table 4.19:

  P(BP = high | diabetes = 0) = 4/11 = 0.36

  P(weight = above | diabetes = 0) = 4/11 = 0.36

  P(FH = yes | diabetes = 0) = 4/11 = 0.36

  P(age = 50 | diabetes = 0) = 3/11 = 0.27

- The **P(X | diabetes=0)** can now be calculated:

  = 0.36 x 0.36 x 0.36 x 0.27

  = **0.0126**

CLASSIFICATION PROCESS.

- The final assessment of **P(X| diabetes=0)P(diabetes= 0)** is computed:

$$P(X|\ diabetes = 0)P(diabetes = 0) = 0.0126 \text{ x } 055 = 0.0069$$

- Since **P(X| diabetes=1)P(diabetes=1) > P(X| diabetes=0) P(diabetes=0)**

$$0.081 > 0.0069$$

- The observations **X are assigned to class diabetes=1**.
- A final probability that diabetes=1, given the evidence (X), can be computed as follows:

$$P(diabetes = 1|\ X) = 0.081/(0.081 + 0.0069) = 0:922$$

ADVANTAGES OF NAÏVE BAYES CLASSIFIER

The naive Bayes is a simple classification approach that works surprisingly well particularly with **large data sets** as well as with **larger numbers of independent variables.**

DISADVANTAGES OF NAÏVE BAYES CLASSIFIER

Only categorical variables:

This method is usually applied in situations in which the independent variables and the response variable are categorical.

Requires large data sets:

This method is versatile, but it is particularly effective in building models from large data sets.


TAKE A BREAK.