Naive Bayes Classifier

Dr. Manju Venugopalan
Asst Professor (S. G)
Dept of CSE
Amrita School of Computing, Bengaluru

Bayes' Theorem:

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

- The posterior probability is equal to the conditional probability of event B given A multiplied by the prior probability of A, all divided by the prior probability of B.

Bayes Theorem

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)}$$

- $P(h)$ = prior probability of hypothesis h
- $P(D)$ = prior probability of training data D
- $P(h|D)$ = probability of h given D
- $P(D|h)$ = probability of D given h

Choosing Hypotheses

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)}$$

Generally want the most probable hypothesis given the training data

Maximum a posteriori hypothesis h_{MAP}:

$$\begin{aligned} h_{MAP} &= \arg\max_{h \in H} P(h|D) \\ &= \arg\max_{h \in H} \frac{P(D|h)P(h)}{P(D)} \\ &= \arg\max_{h \in H} P(D|h)P(h) \end{aligned}$$

If assume $P(h_i) = P(h_j)$ then can further simplify, and choose the *Maximum likelihood* (ML) hypothesis

$$h_{ML} = \arg\max_{h_i \in H} P(D|h_i)$$

Bayes Theorem

Does patient have cancer or not?

A patient takes a lab test and the result comes back positive. The test returns a correct positive result in only 98% of the cases in which the disease is actually present, and a correct negative result in only 97% of the cases in which the disease is not present. Furthermore, .008 of the entire population have this cancer.

$$P(cancer) = \qquad P(\neg cancer) =$$
$$P(+|cancer) = \qquad P(-|cancer) =$$
$$P(+|\neg cancer) = \qquad P(-|\neg cancer) =$$

- $P(cancer) = 0.008$ $P(\neg cancer) = 0.992$
- $P(+|cancer) = 0.98$ $P(-|cancer) = 0.02$
- $P(+|\neg cancer) = 0.03$ $P(-|\neg cancer) = 0.97$

Does patient have cancer or not?

- Suppose we now observe a new patient for whom the lab test returns a **positive** result.
- Should we diagnose the patient as having cancer or not?

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)}$$

$$P(cancer|+) = P(+|cancer) * P(cancer) = 0.98 \; * 0.008 = 0.0078$$

$$P(\urcorner cancer|+) = P(+|\urcorner cancer) * P(\urcorner cancer) = 0.03 \; * 0.992 = 0.0298$$

Does patient have cancer or not?

- Suppose we now observe a new patient for whom the lab test returns a **negative** result.
- Should we diagnose the patient as having cancer or not?

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)}$$

$$P(cancer|-) = P(-|cancer) * P(cancer) = 0.02 \ * 0.008 = 0.00016$$

$$P(\neg cancer|-) = P(-|\neg cancer) * P(\neg cancer) = 0.97 \ * 0.992 = 0.96224$$

$$h_{MAP} = \neg cancer$$

Naive Bayes Classifier

Along with decision trees, neural networks, nearest nbr, one of the most practical learning methods.

When to use

- Moderate or large training set available
- Attributes that describe instances are conditionally independent given classification

Successful applications:

- Diagnosis
- Classifying text documents

Naïve Bayes Analysis

- Naive-Bayes (NB) technique is a supervised learning technique that uses probability-theory-based analysis.
- It is a machine-learning technique that computes the probabilities of an instance belonging to each one of many target classes, given the prior probabilities of classification using individual factors.
- Naïve Bayes technique is used often in classifying text documents into one of multiple predefined categories.

Naïve Bayes Analysis

- NB algorithm is easy to understand and works fast.
- It also performs well in multiclass prediction, such as when the target class has multiple options beyond binary yes/no classification.
- NB can perform well even in case of categorical input variables compared to numerical variable(s)

Naïve Bayes Model

- In the abstract, Naive-Bayes is a conditional probability model for classification purposes.
- The goal is to find a way to predict the class variable (Y) using a vector of independent variables i.e., finding the function f: X →Y.
- In probability terms, the goal is to find P(Y|X), i.e., the probability of Y belonging to a certain class X.
- Y is generally assumed to be a categorical variable with two or more discrete values.

Naive Bayes Classifier

Assume target function $f : X \rightarrow V$, where each instance x described by attributes $\langle a_1, a_2 \ldots a_n \rangle$.
Most probable value of $f(x)$ is:

$$v_{MAP} = \operatorname*{argmax}_{v_j \in V} P(v_j | a_1, a_2 \ldots a_n)$$

$$\begin{aligned} v_{MAP} &= \operatorname*{argmax}_{v_j \in V} \frac{P(a_1, a_2 \ldots a_n | v_j) P(v_j)}{P(a_1, a_2 \ldots a_n)} \\ &= \operatorname*{argmax}_{v_j \in V} P(a_1, a_2 \ldots a_n | v_j) P(v_j) \end{aligned}$$

Naive Bayes assumption:

$$P(a_1, a_2 \ldots a_n | v_j) = \prod_i P(a_i | v_j)$$

which gives

Naive Bayes classifier: $v_{NB} = \operatorname*{argmax}_{v_j \in V} P(v_j) \prod_i P(a_i | v_j)$

NAIVE BAYES CLASSIFIER – Example

Day	*Outlook*	*Temperature*	*Humidity*	*Wind*	*PlayTennis*
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

$\langle Outlook = sunny, Temperature = cool, Humidity = high, Wind = strong \rangle$

P < Outlook = sunny, Temp = cool, Humidity = high, Wind = stro

$$V_{NB} = \underset{v_j \in \{yes, no\}}{\arg\max} \; P(v_j) \prod_i P(a_i | v_j)$$

$P(\text{sunny, cool, high, strong} \mid \text{Yes}) = ?$ $\quad V_{NB}(yes)$

$P(\text{sunny, cool, high, strong} \mid \text{No}) = ?$ $\quad V_{NB}(No)$

Find max.

$$V_{NB}(yes) = P(yes)\left[P(sunny|yes) * P(cool|yes) * P(high|yes) * P(strong|yes)\right]$$

$$= \frac{9}{14}\left[\frac{2}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{3}{9}\right] = 0.0053$$

$$V_{NB}(no) = P(no)\left[P(sunny|no) * P(cool|no) * P(high|no) * P(strong|no)\right]$$

$$= \frac{5}{14}\left[\frac{3}{5} \times \frac{1}{5} \times \frac{4}{5} \times \frac{3}{5}\right] = 0.0206$$

Naïve Bayesian Classifier: Training Dataset

Class:
C1:buys_computer = ‘yes’
C2:buys_computer = ‘no’

age	income	student	credit_rating	_comp
<=30	high	no	fair	no
<=30	high	no	excellent	no
31…40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31…40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31…40	medium	no	excellent	yes
31…40	high	yes	fair	yes
>40	medium	no	excellent	no

Data sample
X = (age <=30,Income = medium,Student = yes,Credit_rating = Fair)

Naïve Bayesian Classifier: An Example

$P(C_i)$: P(buys_computer = “yes”) = 9/14 = 0.643
P(buys_computer = “no”) = 5/14= 0.357

Compute $P(X|C_i)$ for each class

P(age = “<=30” | buys_computer = “yes”) = 2/9 = 0.222
P(income = “medium” | buys_computer = “yes”) = 4/9 = 0.444
P(student = “yes” | buys_computer = “yes) = 6/9 = 0.667
P(credit_rating = “fair” | buys_computer = “yes”) = 6/9 = 0.667

P(age = “<= 30” | buys_computer = “no”) = 3/5 = 0.6
P(income = “medium” | buys_computer = “no”) = 2/5 = 0.4
P(student = “yes” | buys_computer = “no”) = 1/5 = 0.2
P(credit_rating = “fair” | buys_computer = “no”) = 2/5 = 0.4

Naïve Bayesian Classifier: An Example

X = (age <= 30 , income = medium, student = yes, credit_rating = fair)

$P(X|C_i)$:

$P(X|buys\_computer = \text{"yes"}) = 0.222 \times 0.444 \times 0.667 \times 0.667 = 0.044$

$P(X|buys\_computer = \text{"no"}) = 0.6 \times 0.4 \times 0.2 \times 0.4 = 0.019$

$P(X|C_i)*P(C_i)$:

$P(X|buys\_computer = \text{"yes"}) * P(buys\_computer = \text{"yes"}) = 0.028$

$P(X|buys\_computer = \text{"no"}) * P(buys\_computer = \text{"no"}) = 0.007$

Therefore, X belongs to class ("buys_computer = yes")

Avoiding the 0-Probability Problem

Naïve Bayesian prediction requires each conditional prob. be non-zero. Otherwise, the predicted prob. will be zero

$$P(X|C_i) = \prod_{k=1}^{n} P(x_k|C_i)$$

Ex. Suppose a dataset with 1000 tuples, income=low (0), income= medium (990), and income = high (10),

Use Laplacian correction (or Laplacian estimator)

- Adding 1 to each case
  - Prob(income = low) = 1/1003
  - Prob(income = medium) = 991/1003
  - Prob(income = high) = 11/1003
- The “corrected” prob. estimates are close to their “uncorrected” counterparts

Naïve Bayesian Classifier: Comments

- Advantages
  - Easy to implement
  - Good results obtained in most of the cases
- Disadvantages
  - Assumption: class conditional independence, therefore loss of accuracy
  - Practically, dependencies exist among variables
    - E.g., hospitals: patients: Profile: age, family history, etc.
    - Symptoms: fever, cough etc., Disease: lung cancer, diabetes, etc.
    - Dependencies among these cannot be modeled by Naïve Bayesian Classifier

Variations in Naive Bayes

- **Gaussian Naive Bayes:**
  - **Description:** Assumes that the features follow a normal (Gaussian) distribution. It is suitable for continuous data where the distribution of each feature is approximately Gaussian.
- **Multinomial Naive Bayes:**
  - **Description:** Assumes that features follow a multinomial distribution. It is typically used for text classification where features are word counts or frequencies.
- **Bernoulli Naive Bayes:**
  - **Description:** Assumes that features follow a Bernoulli distribution, meaning each feature is binary (0 or 1). It is suitable for binary/boolean features.

.

Thank you