# Machine Learning
# CSE - 465

## Lecture - 05

# Outline

- Bayes' Theorem
- Naïve Bayes Classifier
- Testing naïve Bayes classifier with a dataset
- Avoiding zero-probability problem in naïve Bayes classifier

# Bayes' Theorem (1/2)

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

- Where A and B are events and P(B) ≠ 0
- P(A|B) is a conditional probability: the likelihood of event A occurring given that B is true
- P(B|A) is also a conditional probability: the likelihood of event B occurring given that A is true

# Bayes' Theorem (2/2)

- P(A) and P(B) are the probabilities of observing A and B respectively; they are known as the marginal probability.

- The **marginal probability** is the probability distribution of the variables contained in the subset. It gives the probabilities of various values of the variables in the subset without reference to the values of the other variables

# Naïve Bayes Classifier (1/3)

- Naive Bayes classifiers are a collection of classification algorithms based on **Bayes' Theorem**
- It is not a single algorithm but a family of algorithms where all of them share a common principle, i.e. every pair of features being classified is independent of each other
- All naive Bayes classifiers assume that the value of a particular feature is independent of the value of any other feature, given the class variable

# Naïve Bayes Classifier (2/3)

- For example, a fruit may be considered to be an apple if it is red, round, and about 10 cm in diameter
- A naive Bayes classifier considers each of these features to contribute independently to the probability that this fruit is an apple, regardless of any possible correlations between the color, roundness, and diameter features
- In some special cases naive Bayes classifiers can be trained very efficiently in a supervised learning setting
- Parameter estimation for naive Bayes models uses the method of maximum likelihood

# Naïve Bayes Classifier (3/3)

- Let D be a training set of tuples and their associated class labels, and each tuple is represented by an attribute vector $\mathbf{X} = (x_1, x_2, \ldots, x_n)$
- Suppose there are *m* classes $C_1, C_2, \ldots, C_m$.
- Classification is to derive the maximum posteriori, i.e., the maximal $P(C_i|\mathbf{X})$
- This can be derived from Bayes' theorem

$$P(C_i|\mathbf{X}) = \frac{P(\mathbf{X}|C_i)P(C_i)}{P(\mathbf{X})}$$

- Since P(X) is constant for all classes, only the following needs to be maximized

$$P(C_i|\mathbf{X}) = P(\mathbf{X}|C_i)P(C_i)$$

# 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(age = “<= 30” | buys_computer = “no”) = 3/5 = 0.6
  P(income = “medium” | buys_computer = “yes”) = 4/9 = 0.444
  P(income = “medium” | buys_computer = “no”) = 2/5 = 0.4
  P(student = “yes” | buys_computer = “yes) = 6/9 = 0.667
  P(student = “yes” | buys_computer = “no”) = 1/5 = 0.2
  P(credit_rating = “fair” | buys_computer = “yes”) = 6/9 = 0.667
  P(credit_rating = “fair” | buys_computer = “no”) = 2/5 = 0.4
- **X = (age <= 30 , income = medium, student = yes, credit_rating = fair)**

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

P(X|buys_computer = “yes”) = 0.222 x 0.444 x 0.667 x 0.667 = 0.044

P(X|buys_computer = “no”) = 0.6 x 0.4 x 0.2 x 0.4 = 0.019

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

P(X|buys_computer = “yes”) * P(buys_computer = “yes”) = 0.028

P(X|buys_computer = “no”) * P(buys_computer = “no”) = 0.007

**So, X belongs to class (“buys_computer = yes”)**

# Avoiding the Zero-Probability Problem

- Naïve Bayesian prediction requires each conditional probability be **non-zero**. Otherwise, the predicted probability 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*

    P(income = low) = 1/1003

    P(income = medium) = 991/1003

    P(income = high) = 11/1003
  - The “corrected” prob. estimates are close to their “uncorrected” counterparts

# Thank You