SUMMARY OF NAIVE BAYES

SYRACUSE UNIVERSITY
School of Information Studies

BUILD A NAIVE BAYES CLASSIFIER

naive Bayes Classifier:

Tid	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

P(refund=no) = 7/10
P(refund=yes) = 3/10

Prior probability

P(Refund=Yes|No) = 3/7
P(Refund=No|No) = 4/7
P(Refund=Yes|Yes) = 0
P(Refund=No|Yes) = 1
P(Marital Status=Single|No) = 2/7
P(Marital Status=Divorced|No)=1/7
P(Marital Status=Married|No) = 4/7
P(Marital Status=Single|Yes) = 2/3
P(Marital Status=Divorced|Yes)=1/3
P(Marital Status=Married|Yes) = 0

For taxable income:
If class=No: sample mean=110
sample variance=2975
If class=Yes: sample mean=90
sample variance=25

Conditional probability

USE NAIVE BAYES CLASSIFIER FOR PREDICTION

Given a test record, calculate posterior probabilities, and choose decision with maximum posterior probabilities.

$$X = (\text{Refund} = \text{No}, \text{Married}, \text{Income} = 120\text{K})$$

$$\begin{aligned} P(X|\text{Class=No}) &= P(\text{Refund=No}|\text{Class=No}) \\ &\quad \times P(\text{Married}|\text{Class=No}) \\ &\quad \times P(\text{Income=120K}|\text{Class=No}) \\ &= 4/7 \times 4/7 \times 0.0072 = 0.0024 \end{aligned}$$

$$P(\text{Class=No}|X) = P(X|\text{Class=No}) \text{ x } P(\text{Class=No}) \,/\, P(X)$$

USE NAIVE BAYES CLASSIFIER FOR PREDICTION

Given a test record:

$$X = (\text{Refund} = \text{No}, \text{Married}, \text{Income} = 120\text{K})$$

$$\begin{aligned} P(X|Class=No) &= P(Refund=No|Class=No) \\ &\quad \times P(Married|Class=No) \\ &\quad \times P(Income=120K|Class=No) \\ &= 4/7 \times 4/7 \times 0.0072 = 0.0024 \end{aligned}$$

P(Class=No|X) =P(X|Class=No) x P(Class=No) / P(X)

$$\begin{aligned} P(X|Class=Yes) &= P(Refund=No|Class=Yes) \\ &\quad \times P(Married|Class=Yes) \\ &\quad \times P(Income=120K|Class=Yes) \\ &= 1 \times 0 \times 1.2 \times 10^{-9} = 0 \end{aligned}$$

P(Class=Yes|X) =P(X|Class=Yes) x P(Class=Yes) / P(X)

Don't forget smoothing!

FEATURES OF BAYESIAN LEARNING METHODS

Each observed training example can incrementally decrease or increase the estimated probability that a hypothesis is correct, and therefore the algorithm is robust to inconsistent examples.

Prior knowledge can be combined with observed data to determine the final probability of a hypothesis.

E.g., an unbalanced coin has 60% chance heads, 40% tails.

Bayesian methods provide probabilistic predictions.

E.g., “This pneumonia patient has a 93% chance of complete recovery.”

CHALLENGE OF BAYESIAN METHODS

Practical difficulty

- Requires initial knowledge of many probabilities
- Estimates the probabilities when they are unknown
- May need to assume normal distribution for continuous variables

Significant cost to compute all probabilities

- Specialized assumptions to reduce the computational cost
  - E.g., naive Bayes is fast.
- Independence assumption may not hold for some attributes
  - Use other techniques, such as Bayesian belief networks (BBN).

Domingos, P., & Pazzani, M. (1997). On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning, 29*, 103–30.

Mitchell, T. (1990). *Machine learning*. New York: McGraw-Hill.