VICTORIA UNIVERSITY OF WELLINGTON
Te Whare Wananga o te Upoko o te Ika a Maui

School of Engineering and Computer Science

COMP 307 — Lecture 14

Uncertainty and Probability 2

Bayes Rules and Classification by Naive Bayes

Dr Bing Xue and Prof. Mengjie Zhang

bing.xue@ecs.vuw.ac.nz and mengjie.zhang@ecs.vuw.ac.nz

Outline

- Rules from last lecture
- Bayes Rules
- Naive Bayes
  - Assumption
  - Deal with zero count
- Summary

Last Lecture

- Product Rule:
  P(X,Y)=P(X)*P(Y|X)
- Sum Rule:
  $P(X)=\sum_y P(X, Y)$
- Normalisation:

  $\sum_x P(X)=1$

  $\sum_x P(X/Y)=1$
- Independence
  - ↔ P(X|Y) = P(X)
  - ↔ P(X, Y) = P(X) * P(Y)

Bayes Rules

- P(A,B) = P(A|B) P(B)
- We can also get:
  P(A,B)= P(B|A) P(A)
- Bayes Rules:

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

- More variables

$$P(Y|X_1,\ldots,X_n) = \frac{P(X_1,\ldots,X_n|Y)P(Y)}{P(X_1,\ldots,X_n)}$$

Example (Training) Data Set

	Job	Deposit	Family	Class
A	true	low	single	Approve
B	true	low	couple	Approve
C	true	low	single	Approve
D	true	high	single	Approve
E	false	high	couple	Approve
1	true	low	couple	Reject
2	false	low	couple	Reject
3	true	low	children	Reject
4	false	low	single	Reject
5	false	high	children	Reject

Naive Bayes: Example Classification Task

- Determine whether to approve a mortgage application, given data/features about the client:
  - Whether they have a job (true or false)
  - The level of their deposit (low or high)
  - Their family status (single, couple[but no kids], children)
- Classification: either Approve or Reject
- Given a set of data about past clients and the classification by the Banks experts
- Construct a classifier that will output the right answer (class) when given a new (unseen) client (instance)

Bayes Rules for Classification

- Very simple probability-based technique
- Computes *P(class*|instance data*)*,
  - Choose the class with the *highest* probability.
- Problem: Hard to measure *P(class|data)*
  - e.g. P(Reject|job=true & dep = high & fam=children)
  - Needs lots of examples of (job=true & dep=high & fam=children)
  - Then count the fraction that are Reject.
- Use Bayes Rule!
- Classification:
  - Given features X1,X2,…,Xn $P(Y|X_1,\ldots,X_n) = \frac{P(X_1,\ldots,X_n|Y)P(Y)}{P(X_1,\ldots,X_n)}$
  - Predict the class label Y

Bayes Rules

- Solution: First use Bayes' Law/Rules, calculate the probability of given instance belong to a class:

$$P(class|data) = \frac{P(data|class) \times P(class)}{P(data)}$$

- For example:

$$P(Reject|\text{job=true \& dep = high \& fam=children})$$

$$= \frac{P(\text{job=true \& dep = high \& fam=children}|Reject) \times P(Reject)}{P(\text{job=true \& dep = high \& fam=children})}$$

P(***Reject***|job=true & dep = high & fam=children)

P(***Accept***|job=true & dep = high & fam=children)

Choose the highest probability

Bayes: Measuring Probabilities

- We can measure P(class)
  - P(Approve) = ?
  - P(Reject) = ?
- We can measure P(feature|class)
  - P (job=true|Approve) = ?
  - P (job=true|Reject) = ?
  - P (dep=high|Approve) =?
- But, how do we measure
  - P(job=true & dep = high & fam=children)
  - P(job=true & dep = high & fam=children|Reject)

Naive Bayes

- Assume that the attributes are conditionally independent (Almost always wrong!) :
  - Assume that all features are independent given the class label Y
- $$P(X_1, \ldots, X_n|Y) = \prod_{i=1}^{n} P(X_i|Y)$$
- P (job=true & dep=high & fam=children|***Reject***)
  = P(job = true|Reject)*P(dep = high|Reject) *p(fam = children|Reject)
- Measure each of these probabilities by counting
  - There is usually enough data for this.
  - Problem: careful about dealing with 0 probabilities!
- Given an instance, use the table to compute probabilities of each class

Computing Probabilities: Counting Occurrences

	Approve	**Reject**
class	5	5
job=true	4	2
job=false	1	3
dep=low	3	4
dep=high	2	1
fam=single	3	1
fam=couple	2	2
fam=children	0	2

(Counting Occurrences)

	Approve	**Reject**
P(class)	5/10	5/10
P(job=true|class)	4/5	2/5
P(job=false|class)	1/5	3/5
P(dep=low|class)	3/5	4/5
P(dep=high|class)	2/5	1/5
P(fam=single|class)	3/5	1/5
P(fam=couple|class)	2/5	2/5
P(fam=children|class)	0/5	2/5

(Computing Probabilities)

Using Naive Bayes Classifier

- Classify a new case: (job=true & dep = high & fam=children)

P(***Reject***|job=true & dep = high & fam=children)

= P(job=true & dep = high & fam=children|*Reject*) ×P(*Reject*) / P(job=true & dep = high & fam=children)

= P (job=true|*Reject*) × P (dep=high|*Reject*) × P (fam=children|*Reject*) ×P(*Reject*) / P(job=true & dep = high & fam=children)

= 0.4×0.2×0.4×0.5 / P(job=true & dep = high & fam=children)

= 0.016 / P(job=true & dep = high & fam=children)

Using Naive Bayes Classifier

- Classify a new case: (job=true & dep = high & fam=children)

P(***Accept***|job=true & dep = high & fam=children)

$$= \frac{P(\text{job=true \& dep = high \& fam=children}|\textit{Accept}) \times P(\textit{Accept})}{P(\text{job=true \& dep = high \& fam=children})}$$

$$= \frac{P(\text{job=true}|\textit{Accept}) \times P(\text{dep=high}|\textit{Accept}) \times P(\text{fam=children}|\textit{Accept}) \times P(\textit{Accept})}{P(\text{job=true \& dep = high \& fam=children})}$$

$$= \frac{0.8 \times 0.4 \times 0 \times 0.5}{P(\text{job=true \& dep = high \& fam=children})}$$

$$= \frac{0}{P(\text{job=true \& dep = high \& fam=children})}$$

Dealing with Zero Counts

- Initialise table to contain small constant, e.g. 1
- This is not quite sound, but reasonable in practice

	Approve	**Reject**
class	6	6
job=true	5	3
job=false	2	4
dep=low	4	5
dep=high	3	2
fam=single	4	2
fam=couple	3	3
fam=children	1	3

(Counting Occurrences)

	Approve	**Reject**
P(class)	6/12	6/12
P(job=true|class)	5/7	3/7
P(job=false|class)	2/7	4/7
P(dep=low|class)	4/7	5/7
P(dep=high|class)	3/7	2/7
P(fam=single|class)	4/8	2/8
P(fam=couple|class)	3/8	3/8
P(fam=children|class)	1/8	3/8

(Computing Probabilities)

Using Naive Bayes Classifier

P(***Reject***|job=true & dep = high & fam=children)

$$= \frac{P(\text{job=true \& dep = high \& fam=children}|\textit{Reject}) \times P(\textit{Reject})}{P(\text{job=true \& dep = high \& fam=children})}$$

$$= \frac{P(\text{job=true}|\textit{Reject}) \times P(\text{dep=high}|\textit{Reject}) \times P(\text{fam=children}|\textit{Reject}) \times P(\text{Reject})}{P(\text{job=true \& dep = high \& fam=children})}$$

$$= \frac{3/7 \times 2/7 \times 3/8 \times 1/2}{????} = \frac{18/784}{????}$$

P(***Accept***|job=true & dep = high & fam=children)

$$= \frac{P(\text{job=true \& dep = high \& fam=children}|\textit{Accept}) \times P(\textit{Accept})}{P(\text{job=true \& dep = high \& fam=children})}$$

$$= \frac{P(\text{job=true}|\textit{Accept}) \times P(\text{dep=high}|\textit{Accept}) \times P(\text{fam=children}|\textit{Accept}) \times P(\textit{Accept})}{P(\text{job=true \& dep = high \& fam=children})}$$

$$= \frac{5/7 \times 3/7 \times 1/8 \times 1/2}{????} = \frac{15/784}{????}$$

Summary

1. Bayes Rules: $P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$ → $P(Y|X_1,\dots,X_n) = \frac{P(X_1,\dots,X_n|Y)P(Y)}{P(X_1,\dots,X_n)}$

2. Classification: If *Y* is class label, *X1 .. Xn* features, the probability of an instance belong to a class is

$$P(Y|X_1,\dots,X_n) = \frac{P(X_1,\dots,X_n|Y)P(Y)}{P(X_1,\dots,X_n)}$$

Too Hard

2. Assume features are conditionally independent: *given Y, X1 .. Xn are independent to each other:*

$$P(X_1,\dots,X_n|Y) = \prod_{i=1}^{n} P(X_i|Y)$$

$$P(class|data) = \frac{P(data|class) \times P(class)}{P(data)}$$

Naive Bayes

Chose the maximum probability