CS522 Advanced Database Systems
Classification: Bayesian Classifiers

Chengyu Sun
California State University, Los Angeles

Probabilistic Relationship between Attributes and Class

- Ten middle-aged, divorced, male borrowers have defaulted on their loans, but would the 11th one default as well?

Bayes Theorem

$$P(A \mid B) = \frac{P(B \mid A) P(A)}{P(B)}$$

- Prior and posterior probabilities
  - P(A) and P(A|B)
  - P(B) and P(B|A)

Bayesian Classification

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

- **X** is a given record with attribute values $(x_1, x_2, ..., x_n)$, and C_i is a class
- $P(C_i|\mathbf{X})$ is the probability of **X** belonging to class C_i given **X**'s attribute values
- We predict that **X** belong to C_i if $P(C_i|X) > P(C_j|X)$ for $j \neq i$

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

- $P(\mathbf{X})$ does not need to be calculated
  - *Why??*
- $P(C_i)$??
- $P(\mathbf{X}|C_i)$??

Sample Data

TID	Home Owner	Marital Status	Annual Income	Defaulted Borrower
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
11	No	Single	120K	??

Naive Bayesian Classification

- $\mathbf{X}=(x_1,x_2,\ldots,x_n)$
- Assume the attribute values are conditionally independent of one another (the *naive* assumption)

$$P(\mathbf{X}\mid C_i)=\prod_{i=1}^{n}P(x_i\mid C_i)$$
$$=P(x_1\mid C_i)\times P(x_2\mid C_i)\times\cdots\times P(x_n\mid C_i)$$

Attribute A_k is Categorical

- $P(x_k \mid C_i)$ is the fraction of number of records in C_i with value x_k for attribute A_k

Attribute A_k is Continuous-valued ...

- Assume A_k follows a Gaussian distribution with a mean μ and standard deviation σ

$$\mu=\frac{1}{N}\sum_{i=1}^{N}x_i \qquad s=\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu)^2}$$

s: sample standard deviation

... Attribute A_k is Continuous-valued

$$g(x,\mu,\sigma)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$P(x_k\mid C_i)=g(x_k,\mu_{c_i},\sigma_{c_i})$$

Naive Bayesian Classification Example ...

- P(Default=N|HO=N,MS=S,AI=120K)
- P(Default=Y|HO=N,MS=S,AI=120K)

... Naive Bayesian Classification Example

- Annual Income, Default=No
  - μ=110, σ=54.54
  - P(AI=120K|No)=0.0072
- Annual Income, Default=Yes
  - μ=90, σ=5
  - P(AI=120K|Yes)=1.2×10^{-9}

Avoid Zero $P(x_k|C_i)$

- A zero $P(x_k|C_i)$ would make the whole $P(\mathbf{X}|C_i)$ zero
- To avoid this problem, add 1 to each count – assuming the training set is sufficiently large, the effect of adding one is negligible
- Example: P(Default=Y|HO=N,MS=M,AI=120K)??

About Naive Bayesian Classification

- The most accurate classification *if the conditional independence assumption holds*
- In practice, some attributes may be correlated
  - E.g. education level and annual income

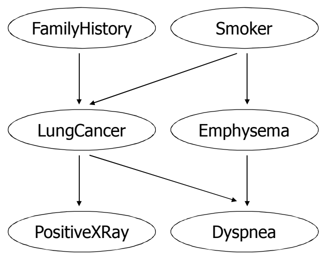
Bayesian Belief Network (BBN)

- A *directed acyclic graph* (dag) encoding the dependencies among a set of variables
- A *conditional probability table* (CPT) for each node given its immediate parent nodes

A BBN Example

CPT for LungCancer

	Yes	No
FH,S	0.8	0.2
FH,!S	0.5	0.5
!FH,S	0.7	0.3
!FH,!S	0.1	0.9



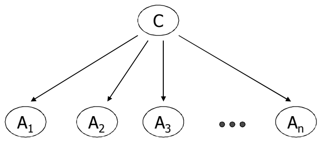
BBN Terminology

- If there is a directed *arc* from X to Y
  - X is a parent of Y
  - Y is a child of X
- If there is a directed *path* from X to Y
  - X is an ancestor of Y
  - Y is a descendent of X

Conditional Independence in BBN

- A node in a Bayesian network is conditionally independent of its non-descendants if its parents are known

Naive Bayesian is a Special Case of BBN



Construct a BBN

- Create the structure of the network
  - From domain knowledge
  - From training data
- Calculate the CPT for each node X
  - P(X) if X does not have any parent
  - P(X|Y) if X has one parent Y
  - $P(X|Y_1,Y_2,\ldots,Y_k)$ if X has multiple parents $\{Y_1,Y_2,\ldots,Y_k\}$

Another BBN Example

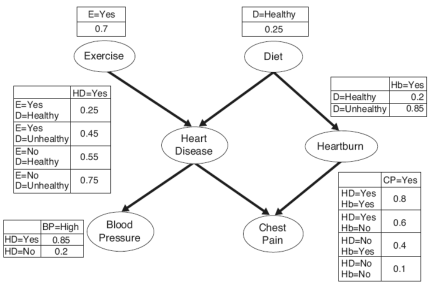


Figure 5.13. A Bayesian belief network for detecting heart disease and heartburn in patients.

Bayesian Classification Examples

- Output node – Heart Disease
- Testing data
  - ()
  - (BP=high)
  - (BP=high,D=Healthy,E=Yes)

Bayesian Classification Examples – 1

$$P(HD=Yes)=\sum_{i=1}^{n}\sum_{j=1}^{m}P(HD=Yes\,|\,E=a_i,D=b_j)P(E=a_i,D=b_j)$$

$$=\sum_{i=1}^{n}\sum_{j=1}^{m}P(HD=Yes\,|\,E=a_i,D=b_j)P(E=a_i)P(D=b_j)$$

$$=0.49$$

Bayesian Classification Examples – 2

$$P(HD=Yes\,|\,BP=High)$$

$$=\frac{P(BP=High\,|\,HD=Yes)P(HD=Yes)}{P(BP=High)}$$

$$=\frac{P(BP=High\,|\,HD=Yes)P(HD=Yes)}{\sum_{i=1}^{n}P(BP=High\,|\,HD=a_i)P(HD=a_i)}$$

$$=0.80$$

Bayesian Classification Examples – 3

$$
\begin{aligned}
&P(HD=Yes \mid BP=High, D=Healthy, E=Yes) \\
&= \frac{P(BP=High \mid HD=Yes, D=Healthy, E=Yes)P(HD=Yes \mid D=Healthy, E=Yes)}{P(BP=High \mid D=Healthy, E=Yes)} \\
&= \frac{P(BP=High \mid HD=Yes)P(HD=Yes \mid D=Healthy, E=Yes)}{\sum_{i=1}^{n} P(BP=High \mid HD=a_i)P(HD=a_i \mid D=Healthy, E=Yes)} \\
&= 0.59
\end{aligned}
$$

About BBN

- Does not assume attribute independence
- Provides a way to encode domain knowledge
  - Robust to model overfitting
- Any node can be used an output node

Bayes Error Rate

- If the relationship between attributes and class is probabilistic, it is impossible to be 100% correct.
- Bayes Error Rate – minimum achievable error rate for a given classification problem

Bayes Error Rate Example ...

- Identify alligators and crocodiles based on their lengths X
- P(X|Crocodile) is Gaussian with μ=15 and σ=2
- P(X|Alligator) is Gaussian with μ=12 and σ=2

...Bayes Error Rate Example...

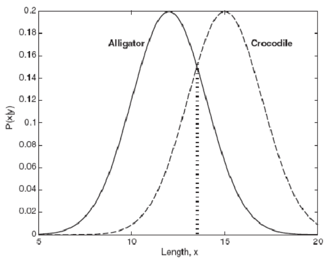


Figure 5.11. Comparing the likelihood functions of a crocodile and an alligator.

... Bayes Error Rate Example

$$Error = \int_0^{\hat{x}} P(X \mid Crocodile)dX + \int_{\hat{x}}^{\infty} P(X \mid Alligator)dX$$

⇩

$$P(X=\hat{x} \mid Crocodile) = P(X=\hat{x} \mid Alligator)$$

⇩

$$\hat{x} = 13.5$$

Readings

- Textbook 6.4.1, 6.4.2, and 6.4.3