CAS CS 506
Lec 13

Naive Bayes

1. Bayes Theorem
   a. P(A|C) = P(C|A) P(A) / P(C)
   b. P(C|A) = P(A n C) / P(A)
   c. P(A|C) = P(A n C) / P(C)
2. Example
   a. Given
      i. Meningitis causes a stiff neck 50% of the time
      ii. Prior probability of any patient having meningitis is 1/50,000
      iii. Prior probability of any patient having a stiff neck is 1/20
   b. If a patient has stiff neck, what is the probability that they have meningitis?

$$P(M|S) = \frac{P(S|M)P(M)}{P(S)} = \frac{.5 \cdot 1/50,000}{1/20} = .0002$$

3. Bayesian Classifiers
   a. Given x = some attribute values
   b. Predict the class C that maximizes P(C | some attribute values)
   c. Example: binary class {yes, no}
      To classify unseen record (material status = “married”, income = 100k)
      i. Compute P(yes | marital status = “married” and income = 100k)
      ii. Compute P(no | martial status = “married” and income = 100k)
      iii. Compare and predict the class that has the highest prob given the attribute values
   d. How do we estimate P(C | some attributes) from the data?
      i.

$$P(C|A_1 \cap A_2 \cap \cdots \cap A_n) = \frac{P(A_1 \cap A_2 \cap \cdots \cap A_n|C)P(C)}{P(A_1 \cap A_2 \cap \cdots \cap A_n)}$$

Maximizing $P(C|A_1A_2\ldots A_n)$ is equivalent to maximizing the numerator on the right

   e. How to estimate $P(A_1\ A_2 \ldots A_n \mid C)\ P(C)$ from the data?
      i. P(C) is easy we can just count how many instances of each class we have
      ii. But $P(A_1\ A_2 \ldots A_n \mid C)$ is tricky because it requires knowing the joint distribution of the attributes

   f. Can we make some assumptions about the attributes in order to simplify the problem?
      i. Assume that A_1 A_2 … A_n are independent
      ii. Then $P(A_1\ A_2 \ldots A_n \mid C) = P(A_1 \mid C) \ldots P(A_n \mid C)$
      iii. Can we estimate $P(A_j \mid C)$ from the data?
         1. Yes! Just count the occurrences of A_j for that class
4. Example
   a.

Refund	Marital Status	Income	Class
Yes	Single	125k	No
No	Married	100k	No
No	Single	70k	No
Yes	Married	120k	No
No	Divorced	90k	Yes
No	Married	60k	No
Yes	Divorced	220k	No
No	Single	85k	Yes
No	Married	75k	No
No	Single	90k	Yes

   b.

Refund	Marital Status	Income	Class
No	Divorced	90k	Yes
No	Single	85k	Yes
No	Single	90k	Yes

P(Marital Status = "Single" | C = Yes) = 2/3

5. Continuous Attributes
   a. Binning / 2 - way or multi-way split
      i. Create new attribute for each bin
      ii. Issue is that these attributes are no longer independent
   b. Pdf Estimation
      i. Assume attribute follows a particular distribution (example: normal)
      ii. Use data to estimate the parameters of the distribution
   c. Assume normal distribution
      i. P(Income = 120k | C = No)

ii.

Refund	Marital Status	Income	Class
Yes	Single	125k	No
No	Married	100k	No
No	Single	70k	No
Yes	Married	120k	No
No	Married	60k	No
Yes	Divorced	220k	No
No	Married	75k	No

1. Sample mean = 110
2. Sample variance = 2975

iii.

$$P(Income = 120|No) = \frac{1}{\sqrt{2\pi}(54.54)} e^{-\frac{(120-110)^2}{2(2975)}} = .0072$$

6. Putting It All Together

a.

Refund	Marital Status	Income	Class
Yes	Single	125k	No
No	Married	100k	No
No	Single	70k	No
Yes	Married	120k	No
No	Divorced	90k	Yes
No	Married	60k	No
Yes	Divorced	220k	No
No	Single	85k	Yes
No	Married	75k	No
No	Single	90k	Yes

Test Record:
X = (Refund = No, Married, Income = 120k)

- P(X | No) = P(Refund = No | No) P(Married | No) P(Income=120k | No) = 4/7 * 4/7 * .0072 = .0024
- P(X | Yes) = P(Refund = No | Yes) P(Married | Yes) P(Income=120k |Yes) = 1 * 0 * 1.2 * 10^{-9} = 0

Since P(X|No)P(No) > P(X|Yes)P(Yes)
=> predict No

7. Limitation
   a. If one of the conditional probabilities is zero, the entire expression becomes zero..
   b. Original estimate of $P(A_i \mid C) = N_{ic} / N_c$
   c. Laplace estimate: $P(A_i \mid C) = (N_{ic} + 1) / (N_c + \text{constant})$
   d. M-estimate: $P(A_i \mid C) = (N_{ic} + mp) / (N_c + m)$
      i. p = prior probability
      ii. m = parameter