APPLICATIONS OF BAYES THEOREM

Shingchern D. You

What to cover

- Iterative method for refining probability. From: http://pansci.asia/archives/66830 (in Chinese)
- Naïve Bayes Classifier. From: https://sebastianraschka.com/Articles/2014_naive_bayes_1.html

Refining probability from evidences

- The love story is in Chinese. I will simplify the story and focus on the computation
- Jane was about to have a blind date with a boy, named Simon
- Jane predicted that P(Simon is a perfect lover) = 0.35 (a priori probability)
- On the date of dating, he came to the Café 30 minutes ahead to occupy a good table with city view

Refining probability from evidences

- Jane determined the followings (based on some reasoning):

  P(Came early in a date | Perfect lover) = 0.9

  P(Came early in a date | Playboy) = 0.7

- With these values, Jane could calcualte P(Simon is a perfect lover | Came early in a date) as

$$\mathrm{P(A|B)} = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^C)P(A^C)} = \frac{0.9 \times 0.35}{0.9 \times 0.35 + 0.7 \times 0.65} = 41\%$$

where A = {Simon is a perfect lover}, B = {Came early in a date}, and A^C = {Simon is a playboy}

Refining probability from evidences

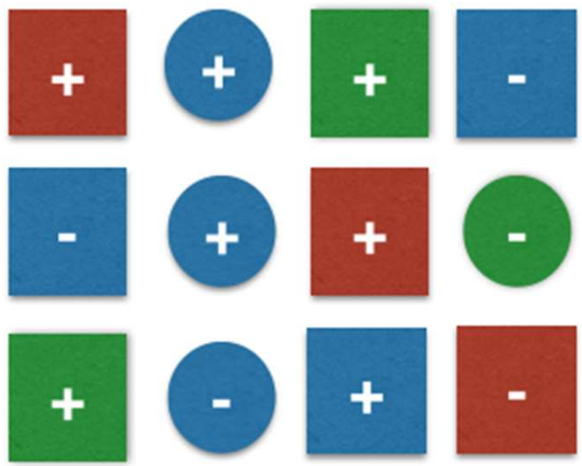
- Jane then took $P(A|B)$ as $P(A)$, i.e., P(Simon is a perfect lover)
- The procedure can be repeated several times to refine $P(A)$
- If you can read Chinese, you may want to read the entire story

Naïve Bayes classifier (1)

- Posteriori probability $P(C \mid \mathbf{x}) = \dfrac{P(C)\,p(\mathbf{x} \mid C)}{p(\mathbf{x})}$
- How to find *P*(C), *P*(***x***|C) and *P*(***x***)?
- Naïve Bayes classifier assumes that all attributes in ***x*** are independent random variables (r.v.)
- Therefore, P(***x***)=P(x_1)P(x_2)…
- We also use empirical probability in the calculation

Naïve Bayes classifier (2)

- Here is an example. The training set is as follows



- $C \in \{+,-\}$, $\mathbf{x} \in \{x_1, x_2\}$, where

  $x_1 \in \{R,G,B\}$, $x_2 \in \{\underline{s}quare, \underline{c}ircle\}$

Naïve Bayes classifier (3)

- To compute the probability, we need a probability model. For this problem, we use the simplest one: We assume the probability of each one-outcome event is equally probable
- If you have a strong belief, other types of model can also be used. However, you may need to estimate the parameters of the model, known as parameteric estimation (to be covered next)

Naïve Bayes classifier (4)

- $P(C=+) = 7/12$, $P(C=-) = 5/12$
- $P(x_1=R) = 3/12$, $P(x_1=G) = 3/12$, $P(x_1=B) = 6/12$
- $P(x_2=s) = 8/12$, $P(x_2=c) = 4/12$
- $P(x_1=R|+) = 2/7$, $P(x_1=G|+) = 2/7$, $P(x_1=B|+) = 3/7$
- $P(x_1=R|-) = 1/5$, $P(x_1=G|-) = 1/5$, $P(x_1=B|-) = 3/5$
- $P(x_2=s|+) = 5/7$, $P(x_2=s|-) = 3/5$
- $P(x_2=c|+) = 2/7$, $P(x_2=c|-) = 2/5$

Naïve Bayes classifier (5)

- Test pattern:  (attributes: B, s)
- $P(\mathbf{x}|+) = P(\text{B}|+)P(\text{s}|+) = (3/7)*(5/7)=15/49$
- $P(\mathbf{x})=P(\text{B})*P(\text{s}) = (6/12)*(8/12)=1/3$
- *Therefore,* $P(+|\mathbf{x}) = P(+)*P(\mathbf{x}|+)/P(\mathbf{x})$
  $=(7/12)*(15/49)/(1/3)=0.5357$
- Similarly, $P(-|\mathbf{x}) = (5/12)*(9/25)/(1/3)=0.45$

Naïve Bayes classifier (6)

- Because $P(+|\mathbf{x}) > P(-|\mathbf{x})$, we let it belong to +, i.e.,

- In the above computation, actually we do not need to compute $P(\mathbf{x})$ because it appears in both $P(+|\mathbf{x})$ and $P(-|\mathbf{x})$ computations

Naïve Bayes classifier (7)

- Read the reference to find out how to deal with unknown attributes (which will lead to probability of zero if not treated)
- In the previous example, we assume that the feature values are of discrete values. Consult the internet to know how to extend it to features with continuous values