Lecture 13: Bayesian learning

CSI 5v93: Introduction to machine learning

Baylor University
Computer Science Department

Dr. Greg Hamerly
http://cs.baylor.edu/~hamerly/

Announcements

- Homework 3 due today
- Homework 4 assigned soon

Questions?

Bayesian learning and naive Bayes

- Bayes' rule and modelling
- naive Bayes
- smoothing
- binning continuous variables
- applications to text
- shaping probabilities

See section 6.6.3 in your book, also handouts from Mitchell.

Bayes' rule

$$\Pr(M|X) = \frac{\Pr(X|M)\Pr(M)}{\Pr(X)}$$

Breaking it apart:

- M –
- X –
- $\Pr(M|X)$ –
- $\Pr(X|M)$ –
- $\Pr(M)$ –
- $\Pr(X)$ –

Bayes' rule

$$\Pr(M|X) = \frac{\Pr(X|M)\Pr(M)}{\Pr(X)}$$

Questions:

- Why are we modelling $\Pr(M|X)$?
- How does each probability affect $\Pr(M|X)$?

Cancer example (from Mitchell's text)

- a test has two outcomes: positive ($\oplus$) or negative ($\ominus$)
- a patient either has cancer (cancer) or does not ($\neg$cancer)

Here are the known probabilities:

$$
\begin{aligned}
P(\text{cancer}) &= .008 & P(\neg\text{cancer}) &= .992 \\
P(\oplus|\text{cancer}) &= .98 & P(\ominus|\text{cancer}) &= .02 \\
P(\oplus|\neg\text{cancer}) &= .03 & P(\ominus|\neg\text{cancer}) &= .97
\end{aligned}
$$

If the test is positive, what is the likely diagnosis: cancer, or not?

Using Bayes' rule for classification

$$\Pr(M|X) = \frac{\Pr(X|M)\Pr(M)}{\Pr(X)}$$

How do we use this to classify an input x?

Different methods, depending on our assumptions:

- MAP – Maximum A Posteriori
- ML – Maximum Likelihood

Naive Bayes classifier

A naive Bayes classifier is based on Bayes' rule.

It's a simple, effective tool for learning from data.

It's called *naive* because of the. . .

Naive (conditional) independence assumption

Assumption: input variable i is independent of input variable j, *given the class*. This is called *conditional independence*.

In probability terms, if input x has features x_i and x_j:

$$\Pr(x_i, x_j | c) = \Pr(x_i | c) \Pr(x_j | c)$$

What does this mean?

Back to Bayes' rule

$$\Pr(M|X) = \frac{\Pr(X|M)\Pr(M)}{\Pr(X)}$$

The naive assumption affects $\Pr(X|M)$, which can be expanded as:

$$\begin{aligned}\Pr(X|M) &= \Pr(X_1|M)\Pr(X_2|M)\cdots\Pr(X_d|M) \\ &= \prod_{i=1}^{d}\Pr(X_i|M)\end{aligned}$$

So instead of having a multivariate model for d-dimensional data, we have d univariate models.

Naive Bayes probabilities

Starting again with Bayes' rule:

$$\Pr(M|X) = \frac{\Pr(X|M)\Pr(M)}{\Pr(X)}$$

and substituting our new definition:

$$\Pr(X|M) = \prod_{i=1}^{d}\Pr(X_i|M)$$

we get the new probability:

$$\Pr(M|X) = \frac{\Pr(M)\prod_{i=1}^{d}\Pr(X_i|M)}{\Pr(X)}$$

Naive Bayes classification

$$\Pr(M|X) = \frac{\Pr(M)\prod_{i=1}^{d}\Pr(X_i|M)}{\Pr(X)}$$

MAP classification:

$$\begin{aligned}
m_{MAP} &= \arg\max_{m \in M} \Pr(M = m|X) \\
&= \arg\max_{m \in M} \frac{\Pr(M = m)\prod_{i=1}^{d}\Pr(X_i|M = m)}{\Pr(X)} \\
&= \arg\max_{m \in M} \Pr(M = m)\prod_{i=1}^{d}\Pr(X_i|M = m)
\end{aligned}$$

2-minute journal

Please write a response to the following on a piece of paper and hand it in immediately. Please make it anonymous (no names). Write about:

- major points you learned today
- areas not understood or requiring clarification