Bayes' Rule

STAT 330 - Iowa State University

Outline

In this lecture students will be introduced to Bayes' rule for "flipping" around conditional probabilities. Bayes' rule is very useful in many applications.

$\mathbb{P}(A|B)$

Bayes' Rule

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B|A) = \mathbb{P}(B)\mathbb{P}(A|B)$$

$$\Rightarrow \boxed{\mathbb{P}(B|A) = \frac{\mathbb{P}(B)\mathbb{P}(A|B)}{\mathbb{P}(A)}}$$

Bayes' Rule

Theorem (Bayes' Rule)

If $B_1, \ldots, B_k$ is a cover or partition of Ω, and A is an event, then

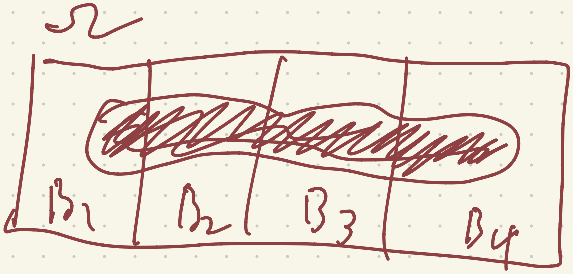
$$\mathbb{P}(B_j|A) = \frac{\mathbb{P}(A|B_j)\mathbb{P}(B_j)}{\sum_{j=1}^{k} \mathbb{P}(A|B_j)P(B_j)}$$

Why?

$$\mathbb{P}(B_j|A) = \frac{\mathbb{P}(A \cap B_j)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B_j)\mathbb{P}(B_j)}{\sum_{j=1}^{k} \mathbb{P}(A|B_j)\mathbb{P}(B_j)}$$

(Definition of conditional probability in numerator, LoTP in denominator)

- Bayes rule $\rightarrow$ way to "flip" conditional probabilities.
- If we know $\mathbb{P}(A|B_j)$ and $\mathbb{P}(B_j)$, then we can obtain $\mathbb{P}(B_j|A)$
- Extremely useful for real world applications!



$\mathbb{P}(B_1), \ldots \mathbb{P}(B_4)$

$\mathbb{P}(A|B_1) \ldots \mathbb{P}(A|B_k)$

$\mathbb{P}(B_j|A)$

---

* $\mathbb{P}(D) = .03$

* $\mathbb{P}(+|D), \quad \mathbb{P}(-|D) \quad \ldots$

$\Rightarrow \quad \underline{\mathbb{P}(D|+)}$

Applying Bayes Rule

Example:

My email is divided into 3 folders: Normal, Important, Spam. From past experience, the probability of emails belonging to these folders is 0.2, 0.1, and 0.7 respectively.

- Out of normal emails, the word "free" occurs with probability 0.01.
- Out of important emails, "free" occurs with probability 0.01.
- Out of spam emails, "free" occurs with probability 0.9.

My spam filter reads an email that contains the word "free". What is the probability that this email is spam?

Applying Bayes Rule Cont.

Define events:

$N =$ email is normal, $I =$ email is important, $S =$ email is spam

$F =$ email contains "free", $\overline{F} =$ email doesn't contain "free"

Given:

$\mathbb{P}(N) = 0.2$, $\mathbb{P}(I) = 0.1$, $\mathbb{P}(S) = 0.7$

$\mathbb{P}(F|N) = 0.01$

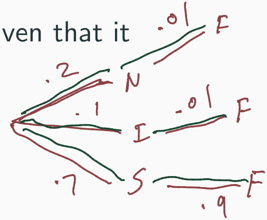
$\mathbb{P}(F|I) = 0.01$

$\mathbb{P}(F|S) = 0.9$

$\mathbb{P}(S|F) =?$ (This is what we want to know)

Applying Bayes Rule Cont.

What is the probability that my email is spam given that it contains the word "free"?



$$\underline{\mathbb{P}(S|F)} = \frac{\mathbb{P}(S \cap F)}{\mathbb{P}(F)}$$

$$= \frac{\mathbb{P}(S)\mathbb{P}(F|S)}{\mathbb{P}(S)\mathbb{P}(F|S) + \mathbb{P}(I)\mathbb{P}(F|I) + \mathbb{P}(N)P(F|N)}$$

$$= \frac{(.7)(.9)}{(.7)(.9) + (.1)(.01) + (.2)(.01)}$$

$$= .995$$

Applying Bayes Rule Cont.

Conceptual understanding

prior probability

- Before knowing anything
  $\rightarrow$ probability that email is spam was $\mathbb{P}(S) = 0.7$.
- After knowing that the email contains the word "free"
  $\rightarrow$ update probability based on this knowledge.

posterior probability

- After knowing the email contains "free"
  $\rightarrow$ probability of the email being spam is $\mathbb{P}(S|F) = 0.995$.
- We could calculate $\mathbb{P}(N|F)$ and $\mathbb{P}(I|F)$ also and *classify* the email to the category with the highest probability.
- In machine learning/statistics, this procedure is called a *naive Bayes classifier*.

Recap

Students should be able to recognize when Bayes' rule is appropriate. They should be able to apply it to practical problems.