Announcements

• Homework 7 is due today

• Exam 2 is next week
Bayes’ Theorem: Suppose that $E$ and $F$ are events from a sample space $S$ such that $p(E) \neq 0$ and $p(F) \neq 0$. Then:

$$p(F|E) = \frac{p(E|F)p(F)}{p(E|F)p(F) + p(E|\overline{F})p(\overline{F})}$$
Bayes’ Theorem: Example

We have two boxes. The first box contains two green balls and seven red balls. The second contains four green balls and three red balls. Bob selects one of the boxes at random. Then he selects a ball from that box at random. If he has a red ball, what is the probability that he selected a ball from the first box.

- Let $E$ be the event that Bob has chosen a red ball and $F$ be the event that Bob has chosen the first box.
- By Bayes’ theorem the probability that Bob has picked the first box is:

$$p(F|E) = \frac{p(E|F)p(F)}{p(E|F)p(F) + p(E|\overline{F})p(\overline{F})}$$
Bayes’ Theorem: Example

We have two boxes. The first box contains two green balls and seven red balls. The second contains four green balls and three red balls. Bob selects one of the boxes at random. Then he selects a ball from that box at random. If he has a red ball, what is the probability that he selected a ball from the first box.

• Let $E$ be the event that Bob has chosen a red ball and $F$ be the event that Bob has chosen the first box.

• By Bayes’ theorem the probability that Bob has picked the first box is:

$$p(F|E) = \frac{p(E|F)p(F)}{p(E|F)p(F) + p(E|\overline{F})p(\overline{F})}$$

$$p(F|E) = \frac{(7/9)(1/2)}{(7/9)(1/2) + (3/7)(1/2)} = \frac{7/18}{38/63} = \frac{49}{76} \approx 0.645.$$
Deriving Bayes’ Theorem

• Recall the definition of the conditional probability \( p(E|F) \):

\[
p(E|F) = \frac{p(E \cap F)}{p(F)}
\]

• From this definition, it follows that:

\[
p(E|F) = \frac{p(E \cap F)}{p(F)} \quad p(F|E) = \frac{p(E \cap F)}{p(E)}
\]
Deriving Bayes’ Theorem

• On the last slide we showed that

\[ p(E|F)p(F) = p(E \cap F) \quad \text{and} \quad p(F|E)p(E) = p(E \cap F) \]

Thus,

\[ p(E|F)p(F) = p(F|E)p(E) \]

Now we can conclude

\[ p(E|F) = \frac{p(F|E)p(E)}{p(F)} \quad p(F|E) = \frac{p(E|F)p(F)}{p(E)} \]
Deriving Bayes’ Theorem

• On the last slide we showed that

\[ p(E|F) = \frac{p(F|E)p(E)}{p(F)} \quad \text{and} \quad p(F|E) = \frac{p(E|F)p(F)}{p(E)} \]

Now we also know (why?)

\[ p(E) = p(E|F)p(F) + p(E|\overline{F})p(\overline{F}) \]

Thus,

\[ p(F|E) = \frac{p(E|F)p(F)}{p(E|F)p(F) + p(E|\overline{F})p(\overline{F})} \]
Bayes’ Theorem: Example 2

Suppose that one person in 100,000 has a particular disease. There is a test for the disease that gives a positive result 99% of the time when given to someone with the disease. When given to someone without the disease, 99.5% of the time it gives a negative result. Find

a) the probability that a person who test positive has the disease.

b) the probability that a person who test negative does not have the disease.

• Should someone who tests positive be worried?
Bayes’ Theorem: Example 2

1 in 100,000 has a particular disease. The test gives a positive result 99% of the time when given to someone with the disease. When given to someone without the disease, 99.5% of the time it gives a negative result. Find a) the probability that a person who test positive has the disease.

Solution: Let $D$ be the event that the person has the disease, and $E$ be the event that this person tests positive.

We need to compute $p(D|E)$.

$$p(D) = 1/100,000 = 0.00001$$  
$$p(D) = 1 - 0.00001 = 0.99999$$

$$p(E|D) = .99$$  
$$p(E|D) = .01$$  
$$p(E|D) = .005$$  
$$p(E|D) = .995$$

$$p(D|E) = \frac{p(E|D)p(D)}{p(E|D)p(D) + p(E|D)p(D)}$$

$$= \frac{(0.99)(0.00001)}{(0.99)(0.00001) + (0.005)(0.99999)}$$

$$\approx 0.002$$

So, don’t worry too much, if your test for this disease comes back positive.
Bayes’ Theorem: Example 2

1 in 100,000 has a particular disease. The test gives a positive result 99% of the time when given to someone with the disease. When given to someone without the disease, 99.5% of the time it gives a negative result. Find

a) the probability that a person who test negative does not have the disease.

Solution:

\[
p(\overline{D}|\overline{E}) = \frac{p(\overline{E}|\overline{D})p(\overline{D})}{p(\overline{E}|\overline{D})p(\overline{D}) + p(\overline{E}|D)p(D)}
\]

\[
= \frac{(0.995)(0.99999)}{(0.995)(0.99999) + (0.01)(0.00001)}
\]

\[
\approx 0.9999999
\]

So, the probability you have the disease if you test negative is

\[
p(D|\overline{E})
\]

\[
\approx 1 - 0.9999999
\]

\[
= 0.0000001.
\]

• So, it is extremely unlikely you have the disease if you test negative.
Generalized Bayes’ Theorem

Generalized Bayes’ Theorem: Suppose that $E$ is an event from a sample space $S$ and that $F_1, F_2, \ldots, F_n$ are mutually exclusive events such that

$$\bigcup_{i=1}^{n} F_i = S.$$ 

Assume that $p(E) \neq 0$ and $p(F_i) \neq 0$ for $i = 1, 2, \ldots, n$. Then

$$p(F_j | E) = \frac{p(E | F_j)p(F_j)}{\sum_{i=1}^{n} p(E | F_i)p(F_i)}.$$
Bayesian Spam Filters

• How do we develop a tool for determining whether an email is likely to be spam?

• If we have an initial set \( B \) of spam messages and set \( G \) of non-spam messages. We can use this information along with Bayes’ law to predict the probability that a new email message is spam.

• We look at a particular word \( w \), and count the number of times that it occurs in \( B \) and in \( G \); \( n_B(w) \) and \( n_G(w) \).
  • Estimated probability that a spam message contains \( w \) is:
    • \( p(w) = n_B(w)/|B| \)

  • Estimated probability that a message that is not spam contains \( w \) is:
    • \( q(w) = n_G(w)/|G| \)
Bayesian Spam Filters

• Let $S$ be the event that the message is spam, and $E$ be the event that the message contains the word $w$.
• Using Bayes’ Rule,

$$p(S|E) = \frac{p(E|S)p(S)}{p(E|S)p(S) + p(E|\overline{S})p(\overline{S})}$$

Assuming that it is equally likely that an arbitrary message is spam and is not spam; i.e., $p(S) = \frac{1}{2}$. We have that,

$$p(S|E) = \frac{p(E|S)}{p(E|S) + p(E|\overline{S})}$$

Note: If we have data on the frequency of spam messages, we can obtain a better estimate.
Bayesian Spam Filters

Let $S$ be the event that the message is spam, and $E$ be the event that the message contains the word $w$. Under our assumption from before, we have that:

$$p(S|E) = \frac{p(E|S)}{p(E|S) + p(E|\bar{S})}$$

Now $p(w)$ and $q(w)$ is our empirical estimate of $p(E | S)$ and $p(E | \bar{S})$

Let $r(w)$ be the estimate of the probability that the message is spam. We can classify the message as spam if $r(w)$ is above a threshold.

$$r(w) = \frac{p(w)}{p(w) + q(w)}$$
Bayesian Spam Filters

**Example:** We find that the word “Rolex” occurs in 250 out of 2000 spam messages and occurs in 5 out of 1000 non-spam messages. Estimate the probability that an incoming message is spam. Suppose our threshold for rejecting the email is 0.9.

**Solution:**

\[
p(\text{Rolex}) = \frac{250}{2000} = 0.125 \quad \text{and} \quad q(\text{Rolex}) = \frac{5}{1000} = 0.005.
\]

\[
r(\text{Rolex}) = \frac{p(\text{Rolex})}{p(\text{Rolex}) + q(\text{Rolex})} = \frac{0.125}{0.125 + 0.005} \approx 0.962
\]

SPAM!
Bayesian Spam Filters

**Example**: We find that the word “Rolex” occurs in 250 out of 2000 spam messages and occurs in 5 out of 1000 non-spam messages. Estimate the probability that an incoming message is spam. Suppose our threshold for rejecting the email is 0.9.

**Solution**: $p(\text{Rolex}) = \frac{250}{2000} = 0.125$ and $q(\text{Rolex}) = \frac{5}{1000} = 0.005$.

$$r(\text{Rolex}) = \frac{p(\text{Rolex})}{p(\text{Rolex}) + q(\text{Rolex})} = \frac{0.125}{0.125 + 0.005} = \frac{0.125}{0.13} \approx 0.962$$

SPAM!
Bayesian Spam Filters

• Accuracy can be improved by considering more than one word as evidence.

• Consider the case where $E_1$ and $E_2$ denote the events that the message contains the words $w_1$ and $w_2$ respectively.

• We make the simplifying assumption that the events are independent. And again we assume that $p(S) = \frac{1}{2}$.

\[
p(S|E_1 \cap E_2) = \frac{p(E_1|S)p(E_2|S)}{p(E_1|S)p(E_2|S) + p(E_1|\bar{S})p(E_2|\bar{S})}
\]

\[
r(w_1, w_2) = \frac{p(w_1)p(w_2)}{p(w_1)p(w_2) + q(w_1)q(w_2)}
\]
Bayesian Spam Filters

Example: We have 2000 spam messages and 1000 non-spam messages. The word “stock” occurs 400 times in the spam messages and 60 times in the non-spam. The word “undervalued” occurs in 200 spam messages and 25 non-spam.

Solution: $p(\text{stock}) = \frac{400}{2000} = .2, \quad q(\text{stock}) = \frac{60}{1000} = .06,$

$p(\text{undervalued}) = \frac{200}{2000} = .1, \quad q(\text{undervalued}) = \frac{25}{1000} = .025$

$$r(\text{stock, undervalued}) = \frac{p(\text{stock})p(\text{undervalued})}{p(\text{stock})p(\text{undervalued}) + q(\text{stock})q(\text{undervalued})}$$

$$= \frac{(0.2)(0.1)}{(0.2)(0.1) + (0.06)(0.025)} \approx 0.930$$
Bayesian Spam Filters

- In general, the more words we consider, the more accurate the spam filter. With the independence assumption if we consider $k$ words:

$$p(S \mid \bigcap_{i=1}^{k} E_i) = \frac{\prod_{i=1}^{k} p(E_i \mid S)}{\prod_{i=1}^{k} p(E_1 \mid S) + \prod_{i=1}^{k} p(E_i \mid \overline{S})}$$

$$r(w_1, w_2, \ldots w_n) = \frac{\prod_{i=1}^{k} p(w_i)}{\prod_{i=1}^{k} p(w_i) + \prod_{i=1}^{k} q(w_i)}$$
Probabilistic Method

Example: Say you have m sets $S_1, S_2, S_3, \ldots, S_m$ such that each set has size $k$ and are subsets of a universe of size $n$. Is it possible to color all of the elements red or blue such that no set is monochromatic (all elements in the set have the same color)

Seems hard to check! Would have to try all possibilities.
Union Bound (Boole’s Inequality)

For any countable set of events $A_1, A_2, \ldots, A_n$

$$p\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} p(A_i)$$
Probabilistic Method

Example: Say you have m sets $S_1, S_2, S_3, \ldots, S_m$ such that each set has size $k$ and are subsets of a universe of size $n$. Is it possible to color all of the elements red or blue such that no set is monochromatic (all elements in the set have the same color)?

Theorem: if $m < 2^{k-1}$ then there is a proper coloring

Proof: Assign each element a random color, red or blue, with equal probability.

$$\Pr(S_i \text{ is monochromatic}) = \Pr(S_i \text{ is all red}) + \Pr(S_i \text{ is blue}) = \frac{1}{2^k} + \frac{1}{2^k} = \frac{1}{2^{k-1}}$$

$$\Pr(\text{any } S_i \text{ is monochromatic}) \leq \sum_i \Pr(S_i \text{ is monochromatic}) \leq m \left(\frac{1}{2^{k-1}}\right)$$

Note that we used the union bound