Math 170E Problem Set 2

Allan Zhang

April 16, 2025

1 Bayes' Rule

1.1 Application of Bayes' rule to medical testing

Suppose there is a new test for a disease, which has a probability of 0.99 of detecting the disease if it is present in the patient (a true positive), but also has a probability of 0.03 of detecting the disease in a person who does not have it (a false positive). Additionally, suppose that the probability that a random person in the population has the disease is 0.05.

(a) What is the probability that a random person who is tested tests positive?

Solution: Let $+$ and $-$ indicate positive and negative tests. Let H and S represent people who do not have the disease and people who do (happy and sad lol)

Then we know that

$$P(+|S) = 0.99 \quad P(+|H) = 0.03 \quad P(S) = 0.05$$

From the law of total probability, we know that

$$P(+) = P(+|S)P(S) + P(+|H)P(H)$$

$$P(+) = 0.99 \cdot 0.05 + 0.03 \cdot 0.95$$

$$P(+) = 0.078$$

(b) What is the probability that a random person has the disease, given they test positive?

Solution: In other words, what is $P(S|+)$

Bayes' theorem states

$$P(S|+) = \frac{P(+|S)}{P(+)}P(S)$$

$$P(S|+) = \frac{0.99}{0.078} \cdot 0.05$$

$$P(S|+) = 0.635$$

1.2 The Naive Bayes test classifier

Suppose you would like to classify texts based on whether they are comedy or news. As training data, you take some collections of news headlines and jokes, and list all the words in all the jokes in a list C and all the words in all the news headlines in a list N.

Each of the lists have 1000 total words. Further, suppose that

- The word 'man' appears 50 times in C and 50 times in N

- The word 'eats' appears 25 times in C and 10 times in N

- The word 'pet' appears 25 times in C and 10 times in N

- The word 'gerbil' appears 25 times in C and 2 times in N

Now you randomly choose one of the two lists C or N with equal probability, then sample (independently, with replacement, in order) four words from that list, and obtain the sample:

$$x = (x1, x2, x3, x4) = (\text{'man', 'eats', 'pet', 'gerbil'})$$

(c) What is the probability $P(C|x)$ that the sample is from C, the words from jokes?

Solution: We know that $P(C)$ and $P(N)$ are both 0.5. Then

$$P(C|x) = \frac{P(x|C)}{P(x)} P(C)$$

We also know that $P(x) = P(x|C)P(C) + P(x|N)P(N)$

$$P(x|C) = P(\text{'man'}|C)P(\text{'eats'}|C)P(\text{'pet'}|C)P(\text{'gerbil'}|C)$$

$$P(x|C) = \frac{50}{1000} \cdot \frac{25}{1000} \cdot \frac{25}{1000} \cdot \frac{25}{1000} = 7.84 \times 10^{-7}$$

$$P(x|N) = P(\text{'man'}|N)P(\text{'eats'}|N)P(\text{'pet'}|N)P(\text{'gerbil'}|N)$$

$$P(x|N) = \frac{50}{1000} \cdot \frac{10}{1000} \cdot \frac{10}{1000} \cdot \frac{2}{1000} = 1 \times 10^{-8}$$

$$P(x) = 0.5(7.84 \times 10^{-7}) + 0.5(1 \times 10^{-8}) = 3.97 \times 10^{-7}$$

$$P(C|x) = \frac{7.84 \times 10^{-7}}{3.97 \times 10^{-7}} \cdot 0.5$$

$$P(C|x) = 0.987$$

(d) What is the probability $P(N|x)$

Solution: We know that $P(N|x) = 1 - P(C|x)$

$$P(N|x) = 1 - 0.987 = 0.013$$

(e) If you later saw this list of words in a new sentence, how would you (naively) decide whether it is a joke or news using the principle of maximum likelihood estimation?

Solution: Since we have calculated $P(C|x)$ to be 0.987 and $P(N|x)$ to be 0.013, we would classify the sentence as a joke

2 The probability space induced by a random variable

(f) Recall the proof from class that for any $A \subset V_X$ we have

$$P_X(A) = \sum_{x \in A} f_X(x)$$

Solution: Remember that

$$P_X(A) = P(\{X \in A\}) = P(\{s \in S | X(s) \in A\}) = P(X^{-1}(A))$$

$$f_X(x) = P(\{X = x\}) = P(\{s \in S | X(s) = x\}) = P(X^{-1}(\{x\}))$$

We also know that $P(A_1 \cup A_2 \cup ... \cup A_n) = P(A_1) + P(A_2) + ... + P(A_n)$, so if we partition the set into single set elements, we can write

$$P_X(A) = \sum_{x \in A} f_X(x)$$

(g) Prove that the function $P_X : \mathcal{P}(V_X) \to [0,1]$ defined a probability measure on V_X, that is, satisfies the definition of a probability space recalled above

Solution: To prove this satisfies the definition of a probability space, we need to show

1. $P(S) = 1$

   By definition,

   $$f_X(x) = P(X^{-1}(\{x\}))$$

   We also proved on the last homework assignment that

   $$S = \bigsqcup_{t \in im(S)} f^{-1}(\{t\})$$

   Then we can state

   $$P_X(A) = P(\bigsqcup_{x \in A} X^{-1}(\{x\})) = \sum_{x \in A} P(X^{-1}(\{x\}))) = \sum_{x \in A} f_X(x)$$

   $$P_X(A) = 1 = \sum_{x \in A} f_X(x)$$

2. $P(A \cup B) = P(A) + P(B)$

   We have already proved that $f^{-1}(A_1 \cup A_2) = f^{-1}(A_1) \cup f^{-1}(A_2)$. Then

   $$P_X(A_1 \cup A_2) = \sum_{x \in A_1 \cup A_2} f_X(x) = \sum_{x \in A_1} f_X(x) + \sum_{x \in A_2} f_X(x)$$

   $$= P_X(A_1) + P_X(A_2)$$

3. $P(A_1 \cup A_2 \cup ...) = P(A_1) + P(A_2) + ...$ for any countable (or finite) list of events $A_1, A_2, ...$ such that $A_i \cap A_j = \emptyset$ for each $i \neq j$

   From our knowledge that the sets $A_1 \cup A_2 \cup ...$ are all disjoint (show in the previous homework set), we can just extend the work we did above to an infinite number of subsets of V_X

   $$P_X(A_1 \cup A_2 \cup ...) = \sum_{x \in A_1 \cup A_2 \cup ...} f_X(x) = \sum_{x \in A_1} f_X(x) + \sum_{x \in A_2} f_X(x) + ... + \sum_{x \in A_n} f_X(x)$$

   $$= P_X(A_1) + P_X(A_2) + ... + P_X(A_n)$$

3   The n coin flips experiment, revisited

Recall the experiment of flipping a coin n independent times, for some number $n \geq 1$, and with probability $p \in (0,1)$ of heads and $q = 1 - p$ of tails, for which the sample space was $S = \{H,T\}^{\times n} = \{s = (s_1, ..., s_n) \mid s_i \in \{H,T\}$ for each $i = 1, ..., n.\}$. Let $X : S \to \mathbb{R}$ be the random variable defined by

$$X(s) = \text{the number of heads flipped in the outcome } s \in S$$

so that for example when $n = 3$ we have $X(TTT) = 0$, $X(TTH) = 1$, and $X(THH) = 2$.

For each of the following questions, provide the general answer as a function of n:

(h) What is V_X and $f_X : V_X \to [0, 1]$ for the random variable X defined above?

Solution:
$$V_X = \{x \in R \mid x = f(s) \text{ for some } s \in S\} = \{0, 1, 2, ..., n\}$$

If we only flip the coin once, since it can either be heads or tails, we could possibly have 0 heads. From there one out, it would be possible to have any number of heads depending on how many times we flip the coin

(i) Compute the expected value of $\mathbb{E}(X)$ for the random variable X defined above

Solution: By definition,
$$\mathbb{E}(X) = \sum_{x \in V_X} x \cdot P(X) = \sum_{x \in V_X} x \cdot f_X(x)$$

In general, the probability of each outcome is

$$p^x (1-p)^{n-x}$$

If we want to take into account the probability that we get exactly x heads, we can add on the binomial coefficient, netting us

$$\frac{n!}{(n-x)x!} \cdot p^x (1-p)^{n-x},$$

which can substitute $f_X(x)$ for us

$$\sum_x^n x \cdot \frac{n!}{(n-x)x!} \cdot p^x (1-p)^{n-x}$$

$$\sum_x^n x \cdot \binom{n}{x} \cdot p^x (1-p)^{n-x}$$

$$x \binom{n}{x} = n \binom{n-1}{x-1}$$

$$\sum_x^n n \cdot \binom{n-1}{x-1} \cdot p^x (1-p)^{n-x}$$

$$n \sum_x^n \binom{n-1}{x-1} \cdot p^x (1-p)^{n-x}$$

$$k = x - 1 \quad n \sum_k^{n-1} \binom{n-1}{k} \cdot p^{k+1} (1-p)^{n-k-1}$$

$$np \sum_k^{n-1} \binom{n-1}{k} \cdot p^k (1-p)^{(n-1)-k}$$

$$np(p + (1-p))^{n-1} = np \cdot 1^{n-1} = np$$

We could simplify the sum, but applying linearity of expectation, which we learned in discussion, is far simpler

$$X_i = 1 \text{ if the } i\text{-th toss is heads} \qquad X_i = 0 \text{ if the } i\text{-th toss is tails}$$

$$\mathbb{E}(X_i) = 0 \cdot (1-p) + 1 \cdot p = p$$

$$\mathbb{E}(X) = \mathbb{E}(X_1 + X_2 + ... + X_n) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + ... + \mathbb{E}(X_n) = np$$

(j) Compute the variance $\sigma^2(X)$ for the random variable X defined above

Solution:

$$\sigma^2(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = E(X^2) - (E(X))^2$$

$$X^2 = X_i X_j | i, j \in n = X_1^2 + X_2^2 + ... + X_n^2 + 2 \sum_{i \leq i < j \leq n} X_i X_j$$

$$\mathbb{E}(X^2) = \mathbb{E}(X_1^2) + \mathbb{E}(X_2^2) + ... + \mathbb{E}(X_n^2) + 2 \sum_{i \leq i < j \leq n} \mathbb{E}(X_i X_j)$$

Since 0^2 and 1^2 remain their original values, we can say

$$\mathbb{E}(X_i^2) = 0 \cdot (1 - p) + 1 \cdot p = p$$

$$\sum_x^n \mathbb{E}(X_i^2) = \sum_x^n \mathbb{E}(X_i) = np$$

$$\mathbb{E}(X_i X_j) = \mathbb{E}(X_i)\mathbb{E}(X_j) = p^2$$

$$2 \cdot \sum_{i \leq i < j \leq n} \mathbb{E}(X_i X_j) = 2 \cdot \binom{n}{2} \cdot p^2 = 2 \cdot \frac{n!}{(n-2)!2!} \cdot p^2 = n(n-1)p^2$$

$$\sigma^2(X) = np + n(n-1)p^2 - n^2 p^2 = np(1 + p(n-1) - np) = np(1-p)$$