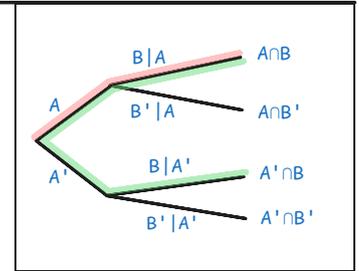


12 - Bayes' Rule

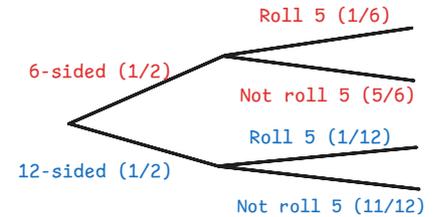
Bayes' rule: $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\text{branch "A then B|A"}}{\text{sum all branches involving B}}$



Example 1. A bag has two dice: a 6-sided and a 12-sided one. I take one out at random and roll a 5. Given this, what's the probability the dice is 6-sided?

$$P(\text{6-sided} \mid \text{roll 5}) = \frac{P(\text{6-sided AND roll 5})}{P(\text{roll 5})}$$

$$= \frac{(1/2)(1/6)}{(1/2)(1/6) + (1/2)(1/12)} = (1/12) / (3/24) = 2/3$$

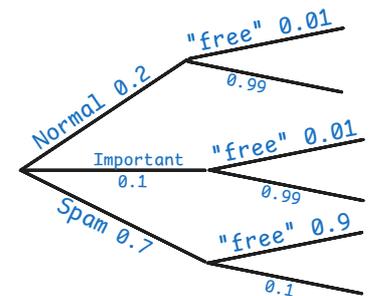


Example 2. (Machine learning application: Naive Bayes classifier)

My emails go into 3 folders: 20% Normal, 10% Important, 70% Spam. In each folder, the word "free" appears in emails with probabilities 0.01, 0.01, and 0.9, respectively. What's the chance an incoming email with the word "free" is spam?

$$P(\text{spam} \mid \text{email says "free"}) = \frac{P(\text{spam AND "free"})}{P(\text{says "free"})}$$

$$= \frac{(0.7)(0.9)}{(0.2)(0.01) + (0.1)(0.01) + (0.7)(0.9)} = \frac{0.63}{0.633} = 0.995$$



This "spam filter" algorithm uses many "spam words" simultaneously with the "naive" assumption that each spam word occurs independently from each other.

Example 3. 10% of Rowans have COVID. A COVID test has:

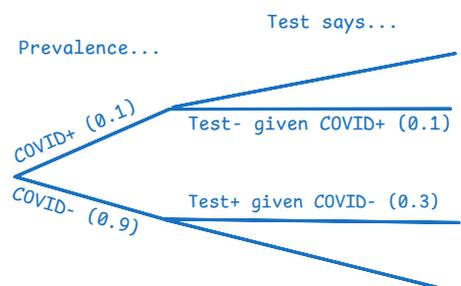
- a false positive rate of 0.3 (test says I have COVID but I don't), and
- a false negative rate of 0.1 (test says I don't have COVID but I do).

If I test positive, what's the chance I have COVID?

Answer: draw tree.

$$P(\text{Is COVID+} \mid \text{Test +}) = \frac{P(\text{Is + and Test+})}{P(\text{Test+})}$$

$$= \frac{(0.1)(0.9)}{(0.1)(0.9) + (0.9)(0.3)} = \frac{0.09}{0.36} = 0.25$$



Moral: Class imbalance in population may result in many false positives or false negatives.

Terminology:

We do binary classification on a population: cancer vs no cancer, criminal vs not criminal, cat vs dog image. An input either has the condition (+) or not (-).

False positive rate (FPR) or Type I error: $\alpha = P(\text{test says } + \mid \text{input is } -)$

False negative rate (FNR) or Type II error rate: $\beta = P(\text{test says } - \mid \text{input is } +)$

Example 4. The function `bool is_prime(int x) { return false; }` passes 95% of 100,000 tests. Find its false positive rate and false negative rate.

$$\text{FPR} = P(\text{algo says prime} \mid p \text{ is not prime}) = 0\%$$

$$\text{FNR} = P(\text{algo says not prime} \mid p \text{ is prime}) = 100\%$$

Moral: Accuracy is bad metric with heavy class imbalance. Instead, use F1-score to avoid extreme FPR and FNR:

$$\frac{2}{F_1} = \frac{1}{\text{TPR}} + \frac{1}{\text{TNR}} = \frac{1}{1 - \alpha} + \frac{1}{1 - \beta}$$

← machine learning/data science concepts, we won't touch this.

1. (Meta, easy)

Facebook has a content team that labels pieces of content on the platform as either spam or not spam. 90% of them are diligent raters and will mark 20% of the content as spam and 80% as non-spam. The remaining 10% are not diligent raters and will mark 0% of the content of spam and 100% as non spam. Assume the pieces of content are labeled independently of one another, for every rater. Given that a rater has labeled 4 pieces of content as good, what is the probability that this rater is a diligent rater?

2. (Microsoft, easy)

Three friends in Seattle each told you it is rainy, and each person has a 1/3 probability of lying. what is the probability that Seattle is raining, assuming that the likelihood of rain on any given day is 0.25?

3. (Meta, easy)

Two coins, one fair (having one side heads and one side tails) and the other unfair (having both sides tails). You pick one at random, flip it five times, and observe that it comes up as tails all five times. What is the probability that you're flipping the unfair coin?

4. (Amazon, easy)

One in 1000 people have a particular disease, and the test for the disease is 98% correct in testing for the disease. on the other hand, the test has a 1% error rate if the person being tested does not have the disease. if someone tests positive, what is the probability that they have the disease?

5. (J.P. Morgan, easy)

A desk has eight drawers. There is a probability of 1/2 that someone placed a letter in one of the desk's 8 drawers and a probability of 1/2 that this person did not place a letter in any of the desk's 8 drawers. You open the first seven drawers and find that they are all empty. What is the probability that the 8th drawer has a letter in it?

(Clarifying assumption: If the person did place a letter in one of the desk's 8 drawers, each drawer is equally likely to contain the letter.)

Answers:
1. 0.79
2. 8/11
3. 0.97
4. 0.0893
5. 1/9