Lecture 9: Naive Bayes' Classifier, More Examples with Conditional Probability.

Anup Rao

January 24, 2018

We see a few more examples using conditional probability, and start talking about random variables.

Example: The Monty Hall Gameshow

In the Monty Hall game show, there are three doors numbered $1, 2, 3$. A uniformly random door has a car behind it. The other two doors each have a goat behind them. The contestants usually prefer cars to goats. The game proceeds as follows:

1. The contestant first picks a door.

2. The host, who knows where the car is, opens a uniformly random door that was not picked by the contestant and has a goat behind it.

3. The contestant then chooses to stay with the door she already picked, or to switch to the other remaining door.

What should the contestant do in the last step?

Suppose the contestant has already picked door number 1. Let us work with the probability space from this point on. Let H_2 be the event that the host opens door number 2, and H_3 be the event that the host opens door number 3. Let C_1, C_2, C_3 be the events that the car is behind door $1, 2$ and 3 respectively. We would like to calculate $p(C_2|H_3)$ and $p(C_1|H_3)$. We can use Bayes' rule to write:

$$p(C_2|H_3) = \frac{p(H_3|C_2) \cdot p(C_2)}{p(H_3)}.$$

$p(H_3|C_2) = 1$, and $p(C_2) = 1/3$. To calculate $p(H_3)$, note that by symmetry we must have $p(H_3) = p(H_2)$ and we have $p(H_3) + p(H_2) = 1$, so $p(H_3) = 1/2$. This gives

$$p(C_2|H_3) = \frac{1/3}{1/2} = \frac{2}{3}.$$

Since given H_3, the car is behind either door 1 or door 2, we must have $p(C_1|H_3) + p(C_2|H_3) = 1$, so we must have $p(C_1|H_3) = 1/3$. This proves that the contestant should always choose to switch to the remaining door if she wishes to maximize her probability of winning!

Example: The Subtleties of Understanding Discrimination

Two different studies aim to determine whether a particular university treats men and women fairly during the admissions process. The first study calculated the acceptance rate for women, and the acceptance rate for men, and found the acceptance rates were the same, and concluded that the system is fair. The second study calculated the acceptance rates within each major, and found that in every single major, the acceptance rate for women was lower than the acceptance rate for men. The conclusion was that the system is unfair. Is it possible that both studies were accurate?

To model this, let us consider the experiment of picking a uniformly random candidate. Let W be the event that the candidate is a woman and M be the event that the candidates is a man. Let A be the event that the candidate is admitted. Then the first study seems to have found

$$p(A|W) = p(A|M).$$

Now, suppose there are two possible majors X, Y that the candidate can apply to. The second study seems to have found:

$$p(A|W, X) < p(A|M, X), p(A|W, Y) < p(A|M, Y).$$

Is this possible?

Consider the probability space shown in figure 1. This satisfies all of the constraints.

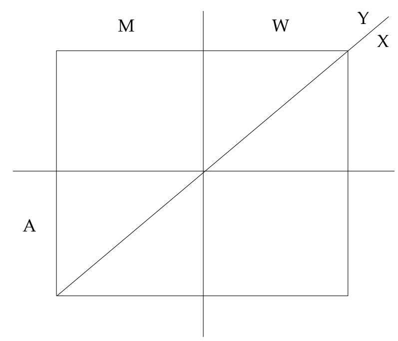

Figure 1: A probability space where all conditions can be met. The region to the left of the vertical line corresponds to men, the region to the right corresponds to women. The region below the horizontal line corresponds to acceptances, and the region below the diagonal corresponds to the major X.