CSM CS70

Interview

Liam Tan

Probabilistic Darts

Definitions:

A. D - The random variable denoting the distance of the 2nd dart from the center

B. A - The event where the 1st dart hits the distance of $\frac{1}{3}$

C. G - The event where the player is good

D. B - The event where the player is bad

We want to find $E[D|A]$

The Good, The Bad, Partitions, and Bayes

There are 2 situations - the player is good or the player is bad.

$E[D|G,A]$ or $E[D|B,A]$

We have to account for both these situations. How can we do that?

We will think about how likely each situation is and accordingly weight the expected value by that probability.

This is called the Law of Total Expectation.

$E[X] = \sum_i E[X|Y_i] * P(Y_i)$

There is a very similar rule called the Total Probability Rule

$P(X) = \sum_i P(X|Y_i) * P(Y_i) = \sum_i P(X \cap Y_i)$

When we can partition a portion of the sample space into disjoint events, we are able to carry out various operations, like calculating probabilities or expectation.

In this case, we split the sample space into the player being good, or the player not being good.

Thus, $1 = P(G) + P(B) = P(G|A) + P(B|A)$

Now, let's think, what is the probability that the player is good, given the information A?

When calculating conditionals, one of the first things we can try is Bayes Rule.

Bayes Rule comes from the definition of conditional probability, as we'll show below.

$P(G|A) = \frac{P(G \cap A)}{P(A)} = \frac{P(A|G)*P(G)}{P(A|G)*P(G)+P(A|B)*P(B)}$

We use total probability to break down $P(A)$ into good and bad partitions, and the multiplication rule to break down the intersection in the numerator.

PDFs and CDFs

So to calculate this probability, we are going to need the probabilities $P(A|G)$ and $P(A|B)$

What do these probabilities represent?

$P(A|G)$ is the probability that we hit distance 1/3 on a circle with radius 1/2

$P(A|B)$ is the probability that we hit distance 1/3 on a circle with radius 1

To do this, we will need the PDF of the random variable R, denoting the distance from the center for dart throw.

Let us derive this with the CDF.

Let's start with the Bad player.

$0 \leq r \leq 1$

$A(r) = \pi r^2$

Here we calculate the CDF with simple geometry? What portion of the circle is radius r or less?

$P(R \leq r) = \frac{A(r)}{A(1)} = r^2$

To find the pdf, we just differentiate and come to $2r$

Similarly, in the same manner, the pdf of the good dart hit is $8r$

To calculate $P(A|G)$ and $P(A|B)$, we simply plug $\frac{1}{3}$ into the pdfs to get $\frac{8}{3}$ and $\frac{1}{3}$, respectively.

Note that the notation isn't entirely correct, as this is a continuous random variable. These are probability densities, not probabilities, that is why we can have numbers greater than 1.

Finally, plugging in our numbers to our equation for $P(G|A)$, we get $\frac{4}{5}$

Since $P(B|A)$ is the complement, we get $\frac{1}{5}$

Conclusions and Expectations

We have an equation here

$E[D|A] = P(G|A) * E[D|G,A] + P(B|A) * E[D|B,A]$

$= \frac{4}{5} * E[D|G,A] + \frac{1}{5} * E[D|B,A]$

How do we calculate these expectations? We use the standard definition of expectation here

$E[D|B] = \int_0^1 y P_y(y) dy = \int_0^1 y * 8y dy = \frac{1}{3}$

$E[D|G] = \int_0^{1/2} y P_y(y) dy = \int_0^{1/2} y * 2y dy = \frac{2}{3}$

Thus the probability is $\frac{4}{5} * \frac{1}{3} + \frac{1}{5} * \frac{2}{3} = \frac{2}{5}$

Uniform Chaos

We first want to think about how to come up with polynomials of degree p-1 or less.

Let us define a polynomial as such: $a_{p-1}x^{p-1} + a_{p-2}x^{p-2} + ... + a_1x^1 + a_0$

We can choose p different coefficients for each a_i and there are p total polynomial coefficients.

Thus there are p^p different polynomials we can make here.

There is a bijection between each polynomial and each string of p numbers from $GF(p)$

Consider each string of numbers as the range of a polynomial.

For example: $3x^2 + 2x + 1$ corresponds to the list $[3, 2, 1]$

Let us work with the p^p lists of numbers.

In this list of list of numbers, I propose that index-wise, at each column, there is a uniform distribution of numbers in each column.

Let's take a look at a simple case where p = 2

1 1

1 0

0 0

0 1

We can see that in each column, there is a uniform distribution of each number in $GF(p)$

How can we formalize this intuition?

Let us consider each string with a fixed number x in the first index. There are p^{p-1} total lists.

By symmetry, for each number in $GF(p)$, there are p^{p-1} total lists where x is fixed in the 1st index.

Thus, each x_i has a $\frac{1}{p}$ probability of being chosen when picked from each column.

This is no different for each index in the column.

Thus, no matter which column is chosen, there is a uniform distribution in the column.

Thus the random polynomial $Q(x)$, no matter which x is chosen, is uniformly distributed across $GF(p)$

Each column is the input to our function Q, and the values in that column are the outputs from the polynomial.

$P(Q(X) = y | X = x) = \frac{1}{p}$

Let us compute $P(Q(x) = y)$ using Total probability

$P(Q(x) = y) = \sum_X P(Q(x) = y | X = x) * P(X = x) = \sum_X \frac{1}{p} * P(X = x) = \frac{1}{p}$

Thus, no matter how X is distributed, the probability of the polynomial being evaluated at an arbitrary value y is always $\frac{1}{p}$. By symmetry, this implies that the random polynomial is uniformly distributed given any distribution X.