Probablity in AI

Justin A. Dang

Almost all machine learning algorithms are predicated on probability. Consider the probability function

$$Pr(x)$$

x is a random variable. When dealing with a random variable, think of it as a collection of data that was randomly observed. Random means it was coincidental as opposed to cherry picked. Note that Pr refers to a collection of functions since data has many representations. Do not assume all probability is measured by one nebulous function.

Types of Data and the Cardinal Rule of Probability

Recall the types of data.

- Discrete
- Continuous

We have different ways of measuring these types of data. With discrete data, the cardinal rule of probability states that

$$\sum_{x \in X} Pr(X = x) = 1$$

and for continuous:

$$\int_{-\infty}^{\infty} Pr(x)dx = 1$$

Consider a discrete function sample. The example for rolling a dice would be:

$$\sum_{1}^{6} Pr(X = x) = 1$$

Area Under the Curve

We know that integrals find the area under the curve. Fundamentally, the integral makes Riemann sum boxes smaller and smaller to get a more accurate area measurement. By setting lower and upper bounds, you can constraint the calculation between two points. So:

$$\lim_{||\Delta|| \to \emptyset} \sum f(x_i) \cdot \Delta x_i$$

$$= \int_{a}^{b} f(x)dx$$

Example of Multivariate Probability

Let x_1 be exam 1 score and x_2 be the average problem set score. How can we determine y, the final grade? This multivariable problem builds a 3D distribution graph. This is represented by multiple integrals with respect to multiple variables.