Fundamentals of Deep Learning

Lec 01: Probability & Information Theory

Vineet Kumar

July 27, 2018

Random Variable

A random variable is just a description of states that are possible for a system (or event). For example:

- Sum of two dice
  - This could take the values $2, 3, \cdots 12$
- Array index of pivot selected in *quick sort*
  - This could take the values $0, 1, ...n - 1$
- A random variable is always accompanied by a *probability* distribution that maps all possible values to a number $\in [0, 1]$
- Question: What is the probability of selecting a random pivot from an array of length n ?

Probability Distribution

For a random variable $\boldsymbol{x}$, probability distribution P should satisfy the following three properties:

- The domain of P is the set of all possible states of a random variable $\boldsymbol{x}$
- $\forall x \in \boldsymbol{x}, 0 \leq P(x) \leq 1$
- $\sum_{x \in \boldsymbol{x}} P(x) = 1$

- Let us say $\boldsymbol{x}$ is the random variable for sum of two dice. Then we can write $P(\boldsymbol{x} = 2) = \frac{1}{36}$

Joint, Marginal & Conditional Probability

Joint Probability

Probability functions can act on more than one random variables. For example $P(\boldsymbol{x} = x, \boldsymbol{y} = y)$, denotes that $\boldsymbol{x} = x$ and $\boldsymbol{y} = y$ occur *simultaneously*.

Marginal Probability

We can reduce a joint probability to a marginal probability by summing over all possible values over the random variable that we want to reduce. This is also known as the **sum rule**

$$P(\boldsymbol{x} = x) = \sum_{y} P(\boldsymbol{x} = x, \boldsymbol{y} = y)$$

Conditional Probability

$$P(\boldsymbol{x} = x | \boldsymbol{y} = y) = \frac{P(\boldsymbol{x} = x, \boldsymbol{y} = y)}{P(\boldsymbol{y} = y)}$$

Chain Rule and Independence

Chain Rule

A joint probability distribution over many random variables can be decomposed using the chain rule as follows)

$$P(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \cdots \boldsymbol{x}^{(n)}) = P(\boldsymbol{x}^{(1)})P(\boldsymbol{x}^{(2)}|\boldsymbol{x}^{(1)})P(\boldsymbol{x}^{(3)}|\boldsymbol{x}^{(2)}, \boldsymbol{x}^{(1)}) \cdots$$

Independence

Two random variables $\boldsymbol{x}$ and $\boldsymbol{y}$ are independent if

$$\forall x \in \boldsymbol{x}, \forall y \in \boldsymbol{y}, P(\boldsymbol{x} = x, \boldsymbol{y} = y) = P(\boldsymbol{x} = x)P(\boldsymbol{y} = y)$$

Expectation

Expectation

Expectation of a random variable $\boldsymbol{x}$ is the average of the values that $\boldsymbol{x}$ can take. Since, each value has a corresponding probability, average is weighted.

$$E[\boldsymbol{x}] = \sum_{x \in \boldsymbol{x}} x \; P(\boldsymbol{x} = x)$$