7.2 Probability Refresher

Probability Refresher

Let us discuss, briefly, some of the rules of probability which we are going to use frequently to derive and explain concepts behind various machine learning algorithms.

Conditional Probability: It gives us formula to calculate probability of occurrence of one event given that some other event has occurred. For example, if we roll an unloaded dice and we know that an even number has turned up, then the probability of that even number to be 2 is given by following equation.

$$P(2|evennumber) = \frac{P(2, evennumber)}{P(evennumber)}$$

In this example the probability of occurrence of 2 is affected by the event of an even number turning up, e.g., if we know that the number turned up was not an even number then P(2)=0, but if we know that an even number has turned up then P(2)=1/3 .

Product Rule(chain rule) : if A and B are two random variables e.g. in a roll of an unloaded dice, A = probability of getting even number , B= probability of getting a number greater than two.

Then, probability of occurrence of both A and B can be given by product rule as follows

$$P(A, B) = P(B|A)P(A) = P(A|B)P(B)$$

Here, P(A,B) is called joint probability distribution of A and B .

The chain rule can be extended to any number of variables.

$$P(A_1, A_2, A_3, A_4) = P(A_1) \times P(A_2|A_1) \times P(A_3|A_1, A_2)(A_4|A_1, A_2, A_3)$$

Independent Random Variables: If occurrence of one does not affect the occurrence of other then we say that these variables are independent of each other.

$$P(B|A) = P(B)$$
$$P(A|B) = P(A)$$
$$P(A, B) = P(B|A)P(A) = P(A) \times P(B)$$

Example: We toss a coin and roll a dice. Let A= event of getting head and B= event of getting an odd number on the dice. We observe that, in this setup, occurrence of A does not help us in predicting the occurrence of B and vice versa.

Conditional Independence: - Given two random variables A and B, we say that they are conditionally independent from each other given that we know a third random variable C such that knowledge of A doesn't provides any additional information about occurrence of B and knowledge of B doesn't provides any information about the occurrence of A. In other words, we need to know only C to figure out about A and B. The same can be written in equations as follows

$$
\begin{aligned}
&P(A, B|C) = P(A|C) \times P(B|C) \\
&or \\
&P(A|B, C) = P(A|C) \\
&P(B|A, C) = P(B|C)
\end{aligned}
$$

A and B are conditionally independent given C.

Example : Let A be the probability of Albert passing an exam , B be the probability of Bob passing the same exam and let C be the probability that student has studied at least 40 hours for the exam . Now if we know C for Albert we can easily find A we do not need information about bob passing the exam.