Task11

Nawar Saeed

Artificial Intelligence for the Web, VT21
DT506A

16 april 2021

This task is about proving that the entropy of a discrete random variable is maximum when all values are equally likely. First, the entropy formula for a discrete random variable x looks like this:

$$H(x) = -\sum_{i \geq 1}^{m} p_i \ log_2 \ p_i \qquad (1)$$

The entropy is a context of a probability distribution. Lets assume that $X = \{1, ..., m\}$, if the elements in X are equally likely, with other words if all elements have the same probabilities, then $p_i = \frac{1}{m}$. This means that the entropy of a discrete random variable is maximal when all values are equally likely "$H(x) = 1$".

Let's take a simple example. Given a random variable x corresponding to the toss of a coin. The probability $P(x = H) = p$, where p = 0.5. Using the equation (1):

$$H(x) = -(0.5 \ log_2(0.5) + 0.5 log_2(0.5)) = -(log_2(0.5)) = 1$$

Let's now try to prove that.

Prove. It is a maximization problem where the objective function is H(x) and the constraints are $\sum_{i \geq 1}^{m} p_i = 1$ and $p_i =\geq 0$.

$$H(x) = -\sum_{i \geq 1}^{m} p_i \ log_2 \ p_i$$

$$s.t \qquad \sum_{i \geq 1}^{m} p_i = 1$$

$$p_i \geq 0$$

Using the Lagrange multiplier method:

$$\ell = (-\sum_{i \geq 1}^{m} p_i \ log_2 \ p_i) + \lambda.(1 - \sum_{i \geq 1}^{m} p_i)$$

Compute the partial derivatives and set them to zero yields:

$$\frac{\partial \ell}{\partial p_i} = -log_2 \ p_i - log_2 \ e - \lambda = 0$$

From this it obtains that $log_2 e + \lambda = -log_2 p_i$ is independent of i and $\sum_{i \geq 1}^{m} p_i = 1$
Finally, the entropy is maximised with $p_i = \frac{1}{m}$, and this can be proved by derivative the λ:

$$\frac{\partial \ell}{\partial \lambda} = 0 \Longrightarrow \sum_{i=1}^{m} p_i = 1 \ yields \ p_i = \frac{1}{m}$$

□