Problem 9

The entropy of a categorical distribution on K values is defined as

$$H(p) = -\sum_{i=1}^{K} p_i log(p_i)$$

Using the method of Lagrange multipliers, find the categorical distribution that has the highest entropy.

Solution

A categorical distribution is a discrete probability distribution over a finite set of K distinct categories or values. We want to find the categorical distribution that maximizes its entropy. Entropy measures the uncertainty or disorder in a probability distribution.

The method of Lagrange Multipliers is used to find the maximum or minimum of a function subject to some constraints. Our constraint in this case is:

$$\sum_{i=1}^{K} p_i = 1$$

λ (lambda) is the Lagrange multiplier associated with the constraint. Our goal is to maximize L with respect to the probabilities $p_1, p_2, ..., p_K$ and λ. To find the maximum, we set the partial derivatives of L with respect to each variable to zero (for each p_i and λ to enforce the constraint):

$$L(p, \lambda) = -\sum_{i=1}^{K} p_i \log(p_i) + \lambda \left(\sum_{i=1}^{K} p_i - 1 \right)$$

$$\frac{\partial L}{\partial p_i} = -\log(p_i) - 1 + \lambda = 0$$

$$\frac{\partial L}{\partial \lambda} = \sum_{i=1}^{K} p_i - 1 = 0$$

From the λ equation:

$$\sum_{i=1}^{K} p_i = 1$$

From the p_i equation:

$$p_i = e^{-(\lambda+1)}$$

Since the probabilities must sum to 1, we have:

$$K \cdot e^{-(\lambda+1)} = 1$$

Solving for λ:

$$\lambda = -1 - \log\left(\frac{1}{K}\right) = \log(K) - 1$$

Substituting for λ:

$$p_i = e^{-(\log(K)-1)} = \frac{1}{K}$$

We can conclude that the categorical distribution that maximizes entropy is uniform distribution. This makes sense, since all the categories are equally likely:

$$p_i = \frac{1}{K}$$

Problem 10 continued on next page...