Homework 6

Exercise 1

1. We have the problem: $L(P_1, P_2, \ldots, P_m) = \sum_{i=1}^{N} \sum_{j=1}^{m} P(j|x_i) \ln P_j$, to which we want to impose a constraint.

The constraint will be $\sum_{j=1}^{m} P_j - 1$, multiplied by a parameter λ, thus $\lambda\left(\sum_{j=1}^{m} P_j - 1\right)$.

Problem is now formulated as:

$$L(P_1, P_2, \ldots, P_m) = \sum_{i=1}^{N} \sum_{j=1}^{m} P(j|x_i) \ln P_j \; + \lambda\left(\sum_{j=1}^{m} P_j - 1\right)$$

2. For each one of P_j's we have that:

$$L(P_1) = \sum_{i=1}^{N} P(1|x_i) \ln P_1 + \lambda P_1 \Rightarrow \frac{\partial L(P_1)}{\partial P_1} = \frac{1}{P_1} \sum_{i=1}^{N} P(1|x_i) + \lambda$$

$$L(P_2) = \sum_{i=1}^{N} P(2|x_i) \ln P_2 + \lambda P_2 \Rightarrow \frac{\partial L(P_2)}{\partial P_2} = \frac{1}{P_2} \sum_{i=1}^{N} P(2|x_i) + \lambda$$

…

$$L(P_m) = \sum_{i=1}^{N} P(m|x_i) \ln P_m + \lambda P_m \Rightarrow \frac{\partial L(P_m)}{\partial P_m} = \frac{1}{P_m} \sum_{i=1}^{N} P(m|x_i) + \lambda$$

By equating each derivative to 0, we have:

$$\frac{\partial L(P_1)}{\partial P_1} = 0 \Rightarrow \frac{1}{P_1} \sum_{i=1}^{N} P(1|x_i) + \lambda \Rightarrow P_1 = -\frac{\sum_{i=1}^{N} P(1|x_i)}{\lambda}$$

$$\frac{\partial L(P_2)}{\partial P_2} = 0 \Rightarrow \frac{1}{P_2} \sum_{i=1}^{N} P(2|x_i) + \lambda \Rightarrow P_2 = -\frac{\sum_{i=1}^{N} P(2|x_i)}{\lambda}$$

…

$$\frac{\partial L(P_m)}{\partial P_m} = 0 \Rightarrow \frac{1}{P_m} \sum_{i=1}^{N} P(m|x_i) + \lambda \Rightarrow P_m = -\frac{\sum_{i=1}^{N} P(m|x_i)}{\lambda}$$

3. Now, let's substitute the P_j's in the constraint equation:

$$\sum_{j=1}^{m} P_j = 1 \Rightarrow -\frac{\sum_{i=1}^{N} P(1|x_i)}{\lambda} - \frac{\sum_{i=1}^{N} P(2|x_i)}{\lambda} \ldots - \frac{\sum_{i=1}^{N} P(m|x_i)}{\lambda} = 1 \Rightarrow -\sum_{j=1}^{m} \sum_{i=1}^{N} \frac{P(j|x_i)}{\lambda} = 1$$

$$\Rightarrow \lambda = -\sum_{j=1}^{m} \sum_{i=1}^{N} P(j|x_i) \Rightarrow \lambda = -\sum_{i=1}^{N} \sum_{j=1}^{m} P(j|x_i) \Rightarrow \lambda = -\sum_{i=1}^{N} 1 \Rightarrow \lambda = -N$$

4. Getting P_j from (2.) and substituting λ from (3.), we get:

$$P_j = -\frac{\sum_{i=1}^{N} P(j|x_i)}{\lambda} \Rightarrow P_j = -\frac{\sum_{i=1}^{N} P(j|x_i)}{-N} \Rightarrow P_j = \frac{\sum_{i=1}^{N} P(j|x_i)}{N}, j = 1,2..,m$$