CMPUT267 Assignment 2 Solutions

Julian Pagcaliwagan

October 7, 2023

Question 1.

1(a)

The values that X_i and Y_i can take are the values within their sample spaces. Let Ω_X and Ω_Y be the sample spaces of X_i and Y_i, respectively. Clearly,

$$\Omega_X = \{\text{sunny}, \text{not sunny}\}$$
$$\Omega_Y = \{\text{free}, \text{not free}\}$$

By assumption, the outcome space for the two random variables X_i and Y_i is the set $\{0, 1\}$. Let us define this outcome space as

$$\chi := \{0, 1\}, \text{ such that}$$

$$\begin{cases} X_i(\text{not sunny}) = 0 \\ X_i(\text{sunny}) = 1 \\ Y_i(\text{not free}) = 0 \\ Y_i(\text{free}) = 1. \end{cases}$$

Ultimately, we are trying to find the parameters that maximize the likelihood of our dataset $\mathcal{D} = \{(x_1, y_1), (x_2, y_2), ...(x_n, y_n)\}$. In other words, we are solving the optimization problem

$$f_{MLE} = \text{argmax } p(\mathcal{D}|\theta)$$

where f_{MLE} is the maximum likelihood and θ is a vector of parameters. We assume $\forall i \in [1, n]$, (x_i, y_i) is conditionally independent given θ, hence

$$f_{MLE} = \text{argmax} \prod_{i=1}^{n} p(x_i, y_i|\theta)$$

At this point, we are concerned with one single distribution $p(x_i, y_i|\theta)$ for some i in $[1, n]$. By assumption, Y_i is dependent on X_i and so the joint distribution of X and Y given the parameters can be expanded to an equivalent form modelling the dependence between the two, as shown below:

$$\begin{aligned} p(x_i, y_i|\theta) &= \frac{p(x_i, y_i, \theta)}{p(\theta)} && \text{Applying Bayes Theorem} \\ &= \frac{p(y_i|x_i, \theta)p(x_i, \theta)}{p(\theta)} && \text{Applying Chain Rule} \\ &= \frac{p(y_i|x_i, \theta)p(x_i|\theta)p(\theta)}{p(\theta)} && \text{Again applying Chain Rule} \\ &= p(y_i|x_i, \theta)p(x_i|\theta) && (*) \end{aligned}$$

One important thing to note is that if the Y_i is dependent on X_i, then value of the prior X_i changes the distribution of $p(y_i|x_i, \theta)$. This means that

$$p(y_i|X_i = 1, \theta) \neq p(y_i|X_i = 0, \theta)$$

or equivalently, that there are different Bernoulli parameters α_0 and α_1 for the distributions $p(y_i|X_i = 0, \theta)$ and $p(y_i|X_i = 1, \theta)$ respectively. We can now define our vector of parameters θ as

$$\theta := (\alpha_0, \alpha_1, \alpha_x)$$

where α_x is the parameter for $p(x_i)$, so our equation $(*)$ is now

$$p(x_i, y_i|\theta) = p(y_i|X_i = 0, \theta)^{1-x_i} p(y_i|X_i = 1, \theta)^{x_i} p(x_i|\theta)$$

(**I used the hint on the announcements to expand the conditional distribution**). Now we can begin to construct the likelihood function f_{MLE} with this information. Our distributions with respect to the parameters in θ can be expanded to their Bernoulli functions:

$$\begin{aligned} p(y_i|X_i = 0, \theta) &= \alpha_0^{y_i}(1-\alpha_0)^{1-y_i} \\ p(y_i|X_i = 1, \theta) &= \alpha_1^{y_i}(1-\alpha_1)^{1-y_i} \\ p(X_i|\theta) &= \alpha_x^{x_i}(1-\alpha_x)^{1-x_i} \end{aligned}$$

Therefore, our likelihood function f_{MLE} can be expanded to

$$f_{MLE} = \text{argmax} \prod_{i=1}^{n} (\alpha_0^{y_i}(1-\alpha_0)^{1-y_i})^{1-x_i} (\alpha_1^{y_i}(1-\alpha_1)^{1-y_i})^{x_i} \alpha_x^{x_i}(1-\alpha_x)^{1-x_i}$$

1(b)

Question 2.

2(a)