Homework 4
Maximum Likelihood - EM Algorithm

CMU 11-755/18-797: Machine Learning for Signal Processing (Fall 2022)

OUT: November 16th, 2022
DUE: **November 30th, 11:59 PM Eastern Time**

START HERE: Instructions

- **Collaboration policy:** Collaboration on solving the homework is allowed, after you have thought about the problems on your own. It is also OK to get clarification (but not solutions) from books or online resources, again after you have thought about the problems on your own. There are two requirements: first, cite your collaborators fully and completely (e.g., "Jane explained to me what is asked in Question 3.4"). Second, write your solution independently: close the book and all of your notes, and send collaborators out of the room, so that the solution comes from you only.
- **Submitting your work:** Assignments should be submitted to Canvas unless explicitly stated otherwise. Please submit all derivation/explanation results as **report_{YourAndrewID}.pdf**. **Each derivation/proof should be completed on a separate page**. Submissions can be handwritten, but should be labeled and clearly legible. Alternatively, submissions are strongly encouraged to be typeset using LaTeX. Please refer to Piazza for detailed instructions for joining Canvas and submitting your homework.
- **Programming**: All programming portions of the assignments should be submitted to Canvas as well. Please submit all codes and output files as **programming_{YourAndrewID}.zip**. We will not be using this for autograding, but rather for plagiarism detection, meaning you may use any language you would like to program.
- **Late submissions**: You have **in total 7 slack days** that you can **freely apply** to any homework. Any homework submitted after running out of slack days will receive zero credit. Please make sure you submit on time.

1 Logistic Regression and Maximum Likelihood Estimation

1.1 Why logistic regression?

In logistic regression we attempt to predict a binary random variable y (i.e. y can be either 0 or 1) as a function of a vector $\mathbf{x}$.

An alternative approach is to consider a linear model, this is

$$y = \boldsymbol{\beta}^\top \mathbf{x} + \varepsilon \tag{1}$$

where ε is a random variable with 0 mean. In this case, we don't consider that $\mathbf{x}$ is a random variable.

1. Show that for any binary random variable y, if $\mathbb{E}(y) = p$, then $\mathrm{Var}(y) = p(1-p)$
2. Show that, using this model, $\mathbb{E}(y) = \boldsymbol{\beta}^\top \mathbf{x}$
3. Using the proposed model, find an expression for $\mathrm{Var}(\varepsilon)$ which depends on $\mathbf{x}$.

Note: In general, we don't use a linear model to predict a binary variable, because, as it is shown in this problem, the variance of the error depends on the input $\mathbf{x}$, which is an assumption that we want to avoid.

1.2 Maximum Likelihood of Logistic Regression

Given a sample $\{(\mathbf{x}_1, y_1), ..., (\mathbf{x}_n, y_n)\}$, using the logistic regression model, the likelihood of is given by

$$\mathcal{L} = \prod_{i=1}^{n} p_i^{y_i} \cdot (1-p_i)^{1-y_i} \tag{2}$$

where $\mathbf{x}_i \in \mathbb{R}^N$, $p_i = \sigma(\boldsymbol{\beta}^\top \mathbf{x}_i)$, and $\sigma(x) = \frac{e^x}{1+e^x}$.

Show that the value of $\boldsymbol{\beta}$ that maximizes $\mathcal{L}$ must satisfy the following equation

$$\sum_{i=1}^{n} y_i \mathbf{x}_i = \sum_{i=1}^{n} \sigma(\boldsymbol{\beta}^\top \mathbf{x}_i)\mathbf{x}_i \tag{3}$$

1.3 Maximum Likelihood Estimator

Let X be a random variable. We say that X is distributed as Uniform(a, b), denoted as $X \sim \mathrm{Uniform}(a, b)$, if its density function is given by

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a, b] \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

Consider an independent identically distributed sample $\{x_1, ..., x_n\}$, such that

$$x_i \sim \mathrm{Uniform}(-\theta, \theta) \tag{5}$$

where θ is a positive unknown parameter.
Find an expression for the Maximum Likelihood Estimator of θ in terms of the sample $\{x_1, ..., x_n\}$.

2 EM, Shift-Invariant Models and Deblurring

2.1 Shift-Invariant Models

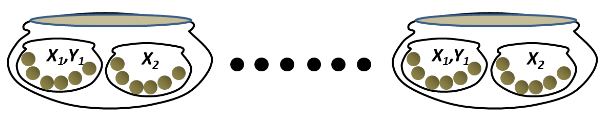
In this problem we will consider shift-invariant mixtures of multi-variate multinomial distributions.

Consider data that have multiple discrete attributes. "Discrete" attributes are attributes that can take only one of a countable set of values. We will consider discrete attributes of a particular kind – integers that have not only a natural rank ordering, but also a definite notion of distance.

Let (X, Y) be the pair of discrete attributes defining any data instance. Since both X and Y are discrete, the probability distribution of (X, Y) is a bi-variate multinomial.

We describe (X, Y) as the outcome of generation by the following process:

The process has at its disposal several urns. Each urn has **two** sub-urns inside it. The first sub-urn represents a bi-variate multinomial: it contains balls, such that each ball has an (X_1, Y_1) value marked on it. The second sub-urn represents a uni-variate multinomial – it contains balls, such that each ball has a X_2 value marked on it.

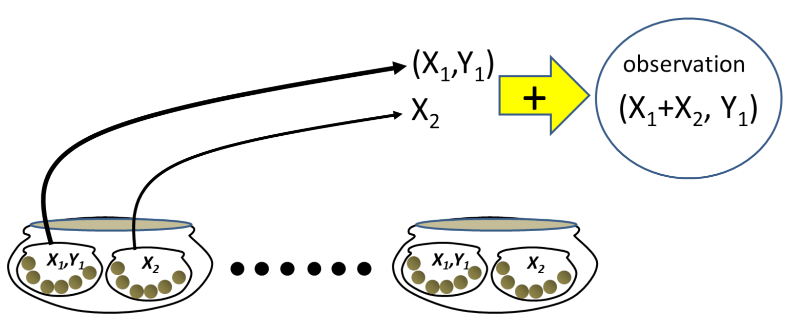


In the following explanation we will use the notation $P_x(X)$ to indicate the probability that the *Random Variable* x takes the value X.

We represent the content of the larger sub-urn within each urn as (x_1, y_1). The smaller sub-urn contains the random variable x_2.

Drawing procedure: At each draw the drawing process performs the following operations.

- It first randomly selects one of the urns according to a probability distribution $P_z(Z)$. Here Z represents the urn selected.
- Then it selects one ball from each of the sub-urns in the selected urn. The probability of balls in the (x_1, y_1) sub-urn of the Z^{th} urn is $P_{x_1,y_1|z}(X_1, Y_1|Z)$. The probability of balls in the x_2 sub-urn of the Z^{th} urn is $P_{x_2|z}(X_2|Z)$. Drawing from these distributions, the process obtains a (X_1, Y_1) pair from the (x_1, y_1) sub-urn, and X_2 from the x_2 sub-urn.
- It finally outputs $(X, Y) = (X_1 + X_2, Y_1)$.



Thus, the final observation is:

$$(X, Y) = (X_1 + X_2, Y_1)$$

Representing the output random variable as (x, y), the probability that it takes a value (X, Y) is given by $P_{x,y}(X, Y)$.

1. Give the expression for $P_{x,y}(X, Y)$ in terms of $P_z(Z)$, $P_{x_1,y_1}(X_1, Y_1|Z)$ and $P_{x_2}(X_2|Z)$. Please, submit your expressions with the proper derivation as part of your report.

2. You are given a histogram of counts $H(X, Y)$ obtained from a large number of observations. $H(X, Y)$ represents the number of times (X, Y) was observed. Give the EM update rules to estimate $P_z(Z)$, $P_{x_1,y_1}(X_1, Y_1|Z)$ and $P_{x_2}(X_2|Z)$. Submit your aswer with the proper derivation as part of your report.

2.2 Application to Deblurring

In this problem we will try to deblur a picture that has become blurry due to a slight left-to-right shake of the camera. You can find the actual picture in `hw4materials_f22/problem2`.

We model the picture as a histogram (the value of any pixel at a position (X, Y), which ranges from 0 to 255, is viewed as the count of "light elements" at that position). We model this distribution as a shift-invariant mixture of one component (i.e. one large urn).

Assuming a very slight 20-pixel strictly-horizontal shake, we model that within the x_2 sub-urn. x_2 can take integer values 0 to 19 (i.e. 20 wide). The x_1 in the (x_1, y_1) sub-urn can range from 0 to (width-of-picture-20). y_1 can take values in the range 0 to (height-of-picture $- 1$).

1. Write a script that takes as input a histogram $H(X, Y)$ and use the EM algorithm to return $P_{x_2}(X_2)$ and $P_{x_1,y_1}(X_1, Y_1)$. Submit your code.

2. Using the image provided, estimate and plot $P_{x_2}(X_2)$ and $P_{x_1,y_1}(X_1, Y_1)$. You will need the solution to problem 2.1 for this problem. Attach your estimation of $P_{x_1,y_1}(X_1, Y_1)$ as an image to your report.