SSU hw03

Jiří Hartvich

January 4, 2023

1 Assignment 1

The task is to solve using MLE the following

$$\frac{1}{m}\sum_{x\in\mathcal{T}^m}\log p(x;s,\eta) \tag{1}$$

For simplicity, I shall for the time being treat x and s as reshaped from the shape $(100, 100)$ to the shape $(100^2,)$ as is custom in Numpy notation (same as the operation `ndarray.flatten()`).

We can thus proceed with our calculations using Equation 2.

$$\log, p(x;s,\eta) = x^T\eta(s) - n_0(s)\log(1+e^{\eta_0}) - n_1(s)\log(1+e^{\eta_1}) \tag{2}$$

1.1 MLE of η

Here we simply take the derivative of Equation 1.

$$\frac{\partial\log p}{\partial\eta} = ... = \frac{1}{m}\left(\sum_{x\in\mathcal{T}^m} x^T[s == 0 \;\; s == 1] - \left[n_0(s)\frac{e^{\eta_0}}{1+e^{\eta_0}} \;\; n_1(s)\frac{e^{\eta_1}}{1+e^{\eta_1}}\right]\right) := 0 \tag{3}$$

Taking all but the x^T elements out of the sum is possible because all of them are orthogonal to x. This gives us the equation

$$x^T(s == k) = n_k(s)\frac{e^{\eta_k}}{1+e^{\eta_k}},\; k = 0, 1 \tag{4}$$

Solving for η_k gives us

$$\eta_k = \log(x_{avg}^T(s == k)) - \log(n(s)_k - x_{avg}^T(s == k)) \tag{5}$$

1.2 MLE of s

Here we solve for s. We shall use the shorthand

$$l_k = \log(1+e^{\eta_k}),\; k = 0, 1.$$

The derivation of s is the following

$$\begin{aligned}\frac{1}{\mathrm{m}}\sum_{\mathrm{x}\in\mathcal{T}^{\mathrm{m}}}\log p(x;s,\eta) = \sum_{i=1}^{N} x_i\eta(s_i) - n_0(s)\,l_0 - n_1(s)\,l_1 = ... = \\ \frac{1}{m}\sum_{i=1}^{N}(1-s_i)(x_i\eta_0 - l_0 + s_i(x_i\eta_1 - l_1) \longrightarrow \max\end{aligned} \tag{6}$$

The same logic of taking the average of x as in the above solution was used here. The solution is then (in Numpy notation)

$$s_i = (x_{avg\,i}\eta_1 - l_1 > x_{avg\,i}\eta_0 - l_0). \tag{7}$$

The algorithm we can use to solve this task is alternating between finding the best η or s while fixing the other of the two.

An appropriate starting η can be calculated from the average image x_{avg} and $s = x_{avg} > 0.5$. Or we can use the tuple $\eta = (a, b)$ where $a < 0\,\&b > 0$.

1.3 Results

The η I obtained is $\eta = [-0.405696470.40651866]$ and the shape s is in Figure 1.

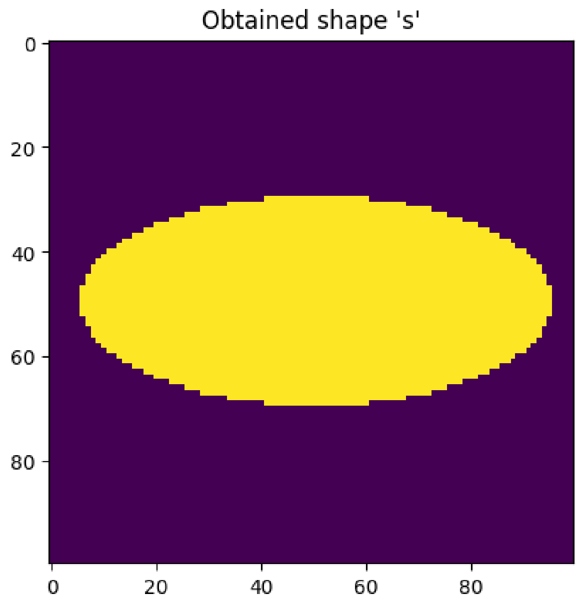

Figure 1: Obtained shape from task 1.

2 Assignment 2

2.1 E step

Let us define the simplified expression:

$$p(x|r; s, \eta) = x^T \eta(T_r s) - K \tag{8}$$

where $K = -n_0(s)\log(1 + e^{\eta_0}) - n_1(s)\log(1 + e^{\eta_1})$, which we are able to do because the expression is orthogonal to r or x.

For the EM algorithm it holds that the lower bound L_B follows

$$\begin{aligned} \arg\max_{\alpha_{\mathrm{x}}(\mathrm{r})} L_B = \arg\max_{\alpha_{\mathrm{x}}(\mathrm{r})} \sum_{x \in \mathcal{T}^m} \sum_{r \in R} [\alpha_x(r) \log p(r|x) - \alpha_x(r) \log \alpha_x(r)] = \\ \arg\max_{\alpha_{\mathrm{x}}(\mathrm{r})} \sum_{x \in \mathcal{T}^m} \sum_{r \in R} [\alpha_x(r)(\log p(x|r) + \log p(r)) - \alpha_x(r) \log \alpha_x(r)] \end{aligned} \tag{9}$$

We maximize this, thus we derivate the expression.

$$\frac{\partial L_B}{\partial \alpha} = \log p(x|r) + \log p(r) - 1 - \log \alpha_x(r) := 0 \tag{10}$$

Plugging in $p(x|r)$ from Equation 8, we get

$$\alpha_x(r) = \exp(x^T \eta(T_r s) - K - 1 + \log p(r)), \sum_{r \in R} \alpha_x(r) = 1 \tag{11}$$

Normalizing so $\sum_{r \in R} \alpha_x(r) = 1$ holds gives us

$$\alpha_x(r) = \mathrm{softmax}_r \left(\langle x \rangle \eta(T_r s) + \log p(r) \right) \tag{12}$$

2.2 M step

2.2.1 π_r

We use Lagrange multipliers.

$$L = \frac{1}{m} \sum_{x \in \mathcal{T}^m} \sum_{r \in R} \alpha_x(r) \log \pi_r - \lambda \left(\sum_{r \in R} \pi_r - 1 \right) \tag{13}$$

$$\frac{\partial L}{\partial \pi_r} = \frac{1}{m} \sum_{x \in \mathcal{T}^m} \sum_{r \in R} \frac{\alpha_x(r)}{\pi_r} - \lambda := 0 \tag{14}$$

$$\frac{\partial L}{\partial \lambda} = \sum_{r \in R} \pi_r - 1 := 0 \tag{15}$$

This gives us the result

$$\pi_r = \frac{1}{m} \sum_{x \in \mathcal{T}^m} \alpha_x(r) \tag{16}$$

2.2.2 η

$$L(x|r; s, \eta) = x^T \eta(T_r s) - n_0(s) \log(1 + e^{\eta_0}) - n_1(s) \log(1 + e^{\eta_1}) \tag{17}$$

Derivating this expression gives us

$$\frac{1}{m} \sum_x \sum_r \alpha_x(r) x^T T_r [s == k] = \eta_k \frac{e^{\eta_k}}{1 + e^{\eta_k}} \tag{18}$$

Here we can clearly see, without knowing anything about dot products in general that $x^T T_r [s == k] = [s == k] T_r^T x$ because it is a scalar. There is added benefit here that we can interpret T_r as a permutation matrix whose inverse is its transpose.

If we set $\psi = \sum_{x \in \mathcal{T}^m} \sum_r \alpha_x(r) T_r^T x$, then the equation now has the same form as in Equation 5 from Assignment 1.

$$\psi^T [s == k] = \eta_k \frac{e^{\eta_k}}{1 + e^{\eta_k}} \tag{19}$$

2.2.3 s

Here we solve for s. We shall use the shorthand

$$l_k = \log(1 + e^{\eta_k}), \ k = 0, 1.$$

$$\frac{1}{m} \sum_{x \in \mathcal{T}^m} \sum_r \alpha_x(r) \left[\log p(x; s, \eta) + \log(\pi_r) \right] \tag{20}$$

The probability π_r is orthogonal to s, so we leave it out. Now, continuing with calculations analogous to those as in equation 6 it follows that

$$\sum_{i=1}^N \sum_r \alpha_x(r) x_i T_r(\eta(s_i)) - n_0(s)\, l_0 - n_1(s)\, l_1 \tag{21}$$

We know from before that $x^T \eta(T_r s) = x^T T_r \eta(s) = \eta(s)^T T_r^T x$. Thus we continue like so

$$\sum_{i=1}^N \sum_r \alpha_x(r) T_r^T(x_i) \eta(s_i) - n_0(s)\, l_0 - n_1(s)\, l_1 \tag{22}$$

$$\sum_{i=1}^N (1 - s_i) \left(\sum_r \alpha_x(r) T_r^T(x_i) \eta_0 - l_0 \right) + s_i \left(\sum_r \alpha_x(r) T_r^T(x_i) \eta_1 - l_1 \right) \longrightarrow \max \tag{23}$$

This gives us

$$\psi = \frac{1}{m} \sum_{x \in \mathcal{T}^m} \sum_r \alpha_x(r) T_r^T x \tag{24}$$

The same logic of taking the average of x as in the above solution was used here. The solution is then (in Numpy notation)

$$s_i = (\psi_i \eta_1 - l_1 > \psi_i \eta_0 - l_0). \tag{25}$$

2.2.4 Stopping criterion

I will stop the EM algorithm when the rate of change of the expression 20 drops below some tolerance threshold, e.g. 1e-5.

2.3 Results

The η I obtained is $\eta = [-0.20144004\ 0.19916591]$, prior probability $\pi = [0.30061295\ 0.20048585\ 0.19999996\ 0.29890124]$ and the shape s is in Figure 2.

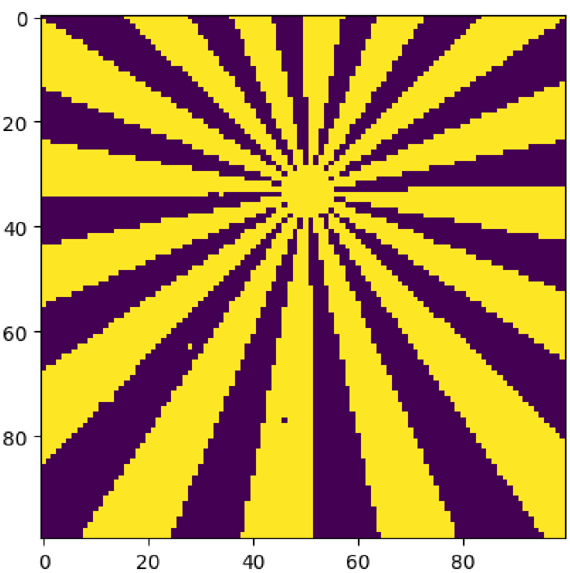

Figure 2: Obtained shape from task 2.