STAT 37601 HW3

Albert Chu

April 2025

1. (a) Given a sample, we can write the likelihood as

$$L(p) = \prod_j^d p_j^{X_j}(1-p_j)^{1-X_j}$$

Therefore, the log likelihood is of the form

$$\sum_j^d (X_j \log p_j + (1-X_i)\log(1-p_j))$$

To find the MLE, we set the score equation to 0,

$$\frac{dl(p)}{dp_j} = \frac{x_j}{p_j} - \frac{1-x_j}{1-p_j} = 0$$

$$\rightarrow x_j - x_j p_j - p_j + x_j p_j = 0 \rightarrow \hat{p_j} = x_j$$

(b) We can rewrite the bayes classifier with the logarithm as

$$c^* = \arg\max_c \log \pi_c + \sum_j^d [x_j \log p_{c,j} + (1-x_j)\log(1-p_{c,j})]$$

Since

$$x_j \log p_{c,j} + (1-x_j)\log(1-p_{c,j}) = x_j \log \frac{p_{c,j}}{1-p_{c,j}} + \log(1-p_{c,j})$$

we can rewrite the classifier as

$$c* = \arg\max_c \log \pi_c + \sum_j^d \left[x_j \log \frac{p_{c,j}}{1-p_{c,j}} + \log(1-p_{c,j})\right]$$

$$= \arg\max_c \sum_j^d x_j \log \frac{p_{c,j}}{1-p_{c,j}} + (\log \pi_c + \sum_j^d \log(1-p_{c,j}))$$

We can then let

$$W_{c,j} = \log \frac{p_{c,j}}{1 - p_{c,j}} \rightarrow W_c = (W_{c,1}, ..., W_{c,d})$$

$$b_c = \log \pi_c + \sum_j^d \log(1 - p_{c,j})$$

which gives us a score for class c of

$$h_c(x) = W_c^T x + b_c$$

(c) i.

$$w_{mi} = P(m|X_i;\theta) = \frac{P(X_i|m;\theta) \cdot P(m;\theta)}{P(X_i;\theta)} = \frac{f_m(X_i;\theta_m) \cdot \pi_m}{\sum_k^M P(X_i|k;\theta) \cdot P(k;\theta)}$$

$$= \frac{\pi_m \cdot f_m(X_i;\theta_m)}{\sum_k^M \pi_k \cdot f_k(X_i;\theta_k)} = \frac{\pi_m \cdot \prod_j^d (p_{j,m})^{X_{ij}} (1 - p_{j,m})^{1-X_{ij}}}{\sum_k^M \pi_k \cdot \prod_j^d p_{j,k}^{X_{ij}} (1 - p_{j,k})^{1-X_{ij}}}$$

ii. For a given data point we can calculate the logarithm of the joint probability of component m and data X_i as

$$\log(\pi_m) + L_{im} = \log(\pi_m) + \sum_j^d X_{ij} \log(p_{j,m}) + (1 - X_{ij}) \log(1 - p_{j,m})$$

We then find the maximum value among these log joint probabilities for the given i value. We can calculate stabilized terms relative to this maximum, M_i,

$$\text{numerator}_{im} = \exp(\text{log joint prob}_{im} - M_i)$$

$$\text{denom}_{ik} = \exp(\text{log joint prob}_{ik} - M_i)$$

The full denominator would be the sum over the k values from 1 to M. Therefore, we can see that since M_i is the max value of the log joint probability, the exponent of the log joint probability minus M_i will always be less than or equal to 0. The numerator for component m will always be less than or equal to 1 and greater than0 as a result.
The denominator is a sum of the denom_{ik} components with each term being less than or equal to 1. Since each of the M terms is at most 1, the sum will be less than or equal to M.

iii. In the M-step, we update, maximizing $\sum_i^n \sum_{m=1}^M w_{mi} \log(\pi_m)$ subject to $\sum_{m=1}^M \pi_m = 1$. Therefore,

$$\pi_m^{new} = \frac{1}{n} \sum_{i=1}^n w_{mi}$$

and the bernoulli parameters as

$$p_{j,m}^{new} = \frac{\sum_{i=1}^{n} w_{mi} \cdot X_{ij}}{\sum_{i=1}^{n} w_{mi}}$$