Assignment 6

Problem 1

In many classification problems, one has the option either of assigning x to class j or, if you are too uncertain, of choosing a "reject" option. If the cost for rejects is less than the cost of falsely classifying the object, it may be the optimal action. Let α_i mean you choose action i, for $i = 1 : C + 1$, where C is the number of classes and $C + 1$ is the reject action. Let $Y = j$ be the true (but unknown) state of nature. Define the loss function as follows

$$\lambda(\alpha_i|Y = j) = \begin{cases} 0 & \text{if } i = j \text{ and } i, j \in \{1, \dots, C\} \\ \lambda_r & \text{if } i = C + 1 \\ \lambda_s, & \text{otherwise} \end{cases}$$

In other words, you incur 0 loss if you correctly classify, you incur λ_r loss (cost) if you choose the reject option, and you incur λ_s loss (cost) if you make a substitution error (misclassification). Both λ_r and λ_s are positive numbers.

1. Show that when we decide to choose a class (and not reject), we always pick the most probable one.
2. Show that the minimum risk is obtained if we decide to pick the most probable class $j_{max} = \arg\max_j p(Y = j|x)$ and if $p(Y = j_{max}|x) \geq 1 - \frac{\lambda_r}{\lambda_s}$; otherwise we decide to reject.
3. Describe qualitatively what happens as λ_r/λ_s is increased from 0 to 1 (i.e., the relative cost of rejection increases).

Problem 2: Bernoulli EM

Consider a unsupervised mixture of Bernoullis model. For each data example i, we have D binary values x_{ij}, we assume there is an unobserved cluster indicator $y_i \in \{1 \dots K\}$, and that $P(x_{ij} \sim \text{Ber}(\mu_{y_{i,j}}))$. In this problem, you will show how to fit this model using EM.

1. Define the evidence lower bound $Q(\theta, \{q_n\})$. Simplify as much as you can.
2. Write an expression that can be used to calculate the cluster responsibility $r_{ik} = P(y_i = k|\theta)$
3. Show that the M step for ML estimation of a mixture of Bernoullis is given by

$$\mu_{kj} = \frac{\sum_i r_{ik} x_{ij}}{\sum_i r_{ik}}$$

Assignment 6

4. Show that the M step for MAP estimation of a mixture of Bernoullis with a $Beta(\alpha, \beta)$ prior is given by

$$\mu_{kj} = \frac{(\sum_i r_{ik} x_{ij}) + \alpha - 1}{(\sum_i r_{ik}) + \alpha + \beta - 2}$$