# MS1b: SDM - Problem Sheet 3

1. (Expectation-Maximization) Assume you are interested in clustering $n$ binary images. Each binary image $\mathbf{x}$ corresponds to a vector of $p$ binary random variables; $p$ being the number of pixels. We adopt a probabilistic approach and model the probability mass function of $\mathbf{x}$ using a finite mixture model

$$f(\mathbf{x}|\theta) = \sum_{k=1}^{K} \pi_k f(\mathbf{x}|\phi_k)$$

where $f(\mathbf{x}|\phi_k)$ is a product of $p$ Bernoulli distributions of parameters $\phi_k = (\phi_k^1, ..., \phi_k^p) \in [0,1]^p$ and $\pi_1, \ldots, \pi_K$ satisfy $\pi_k \geq 0\ \forall k$ and $\sum_{k=1}^{K} \pi_k = 1$. We want to estimate the unknown parameters $\theta = \{\pi_k, \phi_k\}_{k=1}^{K}$ given $\mathbf{x}_{1:n}$ by maximizing the associated likelihood function using the EM algorithm. Establish explicitly the EM update; i.e. the expression of the estimate $\theta^{(t)}$ at iteration $t$ as a function of $\theta^{(t-1)}$.

2. (Decision Theory) Consider two univariate normal distributions $N(\mu, \sigma^2)$ with *known* parameters $\mu_A = 10$ and $\sigma_A = 5$ for class A and $\mu_B = 20$ and $\sigma_B = 5$ for class B. Suppose class A represents the random score $X$ of a medical test of normal patients and class B represents the score of patients with a certain disease. A priori there are 100 times more healthy patients than patients carrying the disease.

   (a) Find the optimal decision rule in terms of misclassification error (0-1 loss) for allocating a new observation $x$ to either class A or B.

   (b) Repeat (a) if the cost of a false negative (allocating a sick patient to group A) is $\theta > 1$ times that of a false positive (allocating a healthy person to group B). Describe how the rule changes as $\theta$ increases. For which value of $\theta$ are 84.1% of all patients with disease correctly classified?

3. (Decision Theory) Suppose there are three equally likely groups of Poisson data, with mean parameters $\lambda_1 = 10$, $\lambda_2 = 15$ and $\lambda_3 = 20$. Show that the optimal rule under 0-1 loss is to allocate to class 1 if $X \leq 12$, to class 2 if $13 \leq X \leq 17$ and to class 3 if $X \geq 18$. Find the class-wise success rates (or probabilities of correct classification) $P(\text{allocate to } i|Y = i)$ for each class i=1,2,3.

4. (Decision Theory) For a given loss function $L$, the risk $R$ of a learner is given by the expected loss

$$R(\hat{Y}) = E(L(Y, \hat{Y})),$$

where $\hat{Y} = \hat{Y}(X)$ is a function of the random predictor variable $X$. Consider a regression problem and the squared error loss

$$L(Y, \hat{Y}) = (Y - \hat{Y})^2.$$

Derive the expression of $\hat{Y} = \hat{Y}(X)$ minimizing the associated risk.