Assignment (EECS6327 F16)

Due: in class on Nov 18, 2016.

You have to work individually. Hand in a hardcopy of your answers before the deadline. No late submission will be accepted. No handwritting is accepted. Direct your queries to Hui Jiang (hj@cse.yorku.ca).

1. (**Missing Features**) Suppose we have three classes in two dimensions with the following underlying distributions:

   - class ω_1: $p(\mathbf{x}|\omega_1) \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
   - class ω_2: $p(\mathbf{x}|\omega_2) \sim \mathcal{N}\left(\binom{1}{1}, \mathbf{I}\right)$
   - class ω_3: $p(\mathbf{x}|\omega_3) \sim \frac{1}{2}\mathcal{N}\left(\binom{0.5}{0.5}, \mathbf{I}\right) + \frac{1}{2}\mathcal{N}\left(\binom{-0.5}{0.5}, \mathbf{I}\right)$

   where $\mathcal{N}(\mu, \Sigma)$ denotes 2-d Gaussian distribution with mean vector μ and covariance matrix Σ, and $\mathbf{I}$ is identity matrix. Assume class prior probabilities $P(\omega_i) = 1/3, i = 1, 2, 3$.

   (a) By explicit calculation of posterior probabilities, classify the feature $\mathbf{x} = \binom{0.3}{0.3}$ based on the MAP decision rule.

   (b) Suppose that for a particular pattern the first feature is missing. Classify $\mathbf{x} = \binom{*}{0.3}$ for minimum probability of error.

   (c) Suppose that for another pattern the second feature is missing. Classify $\mathbf{x} = \binom{0.3}{*}$ for minimum probability of error.

2. (**Maximum Likelihood Estimation**) Assume we have K different classes, i.e. $\omega_1, \omega_2, \cdots, \omega_K$. Each class ω_k $(k = 1, 2, \cdots, K)$ is modeled by a multivariate Gaussian distribution with the mean vector $\boldsymbol{\mu}_k$ and the covariance matrix $\boldsymbol{\Sigma}$, i.e., $p(\mathbf{x} \mid \omega_k) = \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is the common covariance matrix for all K classes. Suppose we have collected N data samples from these K classes, i.e., $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N\}$, and let $\{l_1, l_2, \cdots, l_N\}$ be their labels so that $l_n = k$ means the data sample $\mathbf{x}_n$ comes from the k-th class, ω_k.

Based on the given data set, derive the maximum-likelihood estimates for all model parameters, i.e., all mean vectors $\boldsymbol{\mu}_k$ $(k = 1, 2, \cdots, K)$ and the common covariance matrix $\boldsymbol{\Sigma}$.

3. (**EM algorithm**) Consider a D-dimensional variable $\mathbf{x}$, each of whose dimensions, x_d, is an integer. Suppose the distribution of these variables is described by a mixture of the multinomial distributions so that

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \, p(\mathbf{x}|\boldsymbol{\mu}_k) \propto \sum_{k=1}^{K} \pi_k \prod_{d=1}^{D} \mu_{kd}^{x_d}$$

where the parameter μ_{kd} denotes the probability of d-th dimension in k-th component, subject to $0 \leq \mu_{kd} \leq 1$ $(\forall k, d)$ and $\sum_d \mu_{kd} = 1$ $(\forall k)$.

Given an observed data set $\{\mathbf{x}_n\}$, where $n = 1, \cdots, N$, derive the E and M step equations of the EM algorithm for optimizing the mixing weights π_k $(\sum_k \pi_k = 1)$ and the component parameters μ_{kd} of this distribution by maximum likelihood.

4. (**Bayesian Networks**) Consider three binary random variables $a, b, c \in \{0, 1\}$ having the joint distribution given in Figure 4. Show by direct evaluation that this distribution has the property that a and b are marginally dependent, so that $p(a, b) \neq p(a)p(b)$, but that they become independent when conditioned on c, so that $p(a, b|c) = p(a|c)p(b|c)$.

Based on this observation, draw the corresponding directed graph for a, b, c, and justify it based on the conditional probabilities for all edges.

a	b	c	$p(a,b,c)$
0	0	0	0.192
0	0	1	0.144
0	1	0	0.048
0	1	1	0.216
1	0	0	0.192
1	0	1	0.064
1	1	0	0.048
1	1	1	0.096

Figure 1: The joint distribution over a, b, c.