Advanced Machine Learning
Class Test II Solutions

1. In a course the probability that a student gets a grade "A" is $P(A) = \frac{1}{2}$, a "B" grade is $P(B) = \mu$, a grade "C" is $P(C) = 2\mu$, and a grade"D" is $P(D) = \frac{1}{2} - 3\mu$. We are told that c students get "C" and d students get "D". We do not know how many students got exactly an "A" or exactly a "B". But we do know that h students got either "A" or "B", i.e., $a + b = h$. Our goal is to use the Expectation Maximization algorithm to obtain an estimate of μ. Derive the E-Step and the M-Step. Show your work.

Ans: We can represent the grades by a four valued discrete distribution.

$$p(A) = \frac{1}{2}, \quad p(B) = \mu, \quad p(C) = 2\mu, \quad p(D) = \frac{1}{2} - 3\mu$$

The parameter of the distribution is μ. The latent variable is a and b. Let $\hat{a}, \hat{b}$ represent the expected value of a, b. In the E-Step we keep the distribution parameters constant and estimate the expected value of the latent variables. Thus, we have,

$$\frac{\hat{a}}{a+b+c+d} = \frac{1}{2}, \quad \frac{\hat{b}}{a+b+c+d} = \mu.$$

Combining the above, and, $\hat{a} + \hat{b} = h$, we get -

E-Step:

$$\hat{a} = \frac{\frac{1}{2}}{\frac{1}{2} + \mu} h, \quad \hat{b} = \frac{\mu}{\frac{1}{2} + \mu} h.$$

In the M-Step, given the expected value of the latent variables we compute the maximum likelihood estimate $\hat{\mu}$ of the parameter. This estimate is obtained by considering the probabilities of b, c, d from the given conditions.

M-Step:

$$\hat{\mu} = \frac{h - a + c}{6(h - a + c + d)}.$$

Then we iterate between E and M Steps.

2. Suppose that $p(x)$ is some fixed distribution and that we wish to approximate it using a Gaussian distribution $q(x) = \mathcal{N}(x|\mu, \Sigma)$. By writing down the form of the KL divergence $KL(q||p)$, show that minimization of the KL divergence with respect to μ and Σ leads to the result that μ is given by the expectation of x under $p(x)$ and that Σ is given by its covariance.

Ans: We have, $KL(q(x)||p(x)) = \int q(x) \log q(x)/p(x) = \int q(x) \log q(x) - \log p(x)$. Substitute $q(x) = \mathcal{N}(x|\mu, \Sigma)$, and note that the second term in the expression for KL divergence is independent of μ and Σ.

Now take derivative of the expression of KL divergence wrt μ and Σ and equate them to zero. Also apply the definition of expectation. This leads to the result.