**CSE512 – Assignment 3**

**1. Mistake bound in consistent online learning**

We know that the hypothesis is consistent with the given input data as follows:

: {}

Predictions are made by the algorithm by randomly choosing the hypothesis, the hypothesis being removed when a wrong prediction is made. Thus, after a wrong prediction . Using the realizability assumption, we can assert that is a non-empty set:

By rearranging we can obtain the mistake bound as:

Now, let’s consider that, and , where .

Let, , , t = 1...k.

The algorithm may predict for every t ∈ [k].

The number of mistakes made by the algorithm is .   
Thus, it is proved that the given condition is not a strict inequality.

**2. Stochastic online k-means algorithm**

Ans. We have been given a set of binary variables and K-means objective function .

Now, let’s take derivative of the K-means function G(S,k) wrt , we get:

The Robbins-Monro procedure defines a sequence of successive estimates for, which is as follows:

Here, will be prototype nearest to and the factor of 2 has been absorbed into

Hence, we arrive at a stochastic definition of the K-Means algorithm.

**3. Gaussian mixture model with identical variance.**

Ans. We are given a univariate Gaussians with each component have the same variance

**4. Biased estimation of Gaussian variance.**

Ans. We have to prove that where m is the number of iid random variables.

Let’s assume:

**--------- (a)**

**--------- (b)** (where if i )

Using (a) and (b) we can replace as follows:

From equation **(a)** and **(b)**, we know that and . Thus, we can replace **--------- (c)**

Using the fact that - and from **(c)** we have that

, which completes our proof.

**5. Regularized Maximum Likelihood.**

Ans. We are given a regularized loss function , we need to show that regularized loss is equivalent to the empirical risk if we add 2 new observations to our training set.

So, we begin by adding 1 negative and 1 positive example, denoted by for which corresponding probabilities are - . Thus, we need to minimize the ERM wrt new set. MLE given is:

**---------(a)**

We have been given the equation:

Using the hint provided, we get: **----------(b)**

Combining RHS of **(a)** and **(b)**, we get:

After applying Hoeffding’s inequality, we have:

Hence, with the given confidence parameter , we get:

**6. Bernoulli Mixture**

Ans. Suppose we have a set of d binary variables , where i = (1…d) which follow Bernoulli distribution with parameter :

**----------(a)**

Now, let , from equation (a) we have:

**--------- (b)**

**--------- (c)**

Now,

(multiplying **(a)** and **(b)**)

**--------- (d)**

If we marginalize the equation **(d)** over z, we get: