**Question I: Mistake bound in consistent online learning**

**Solution:**The algorithm maintains a set of , of all the hypotheses which are consistent with the input data, i.e. .  
It picks any hypothesis from and predicts according to this hypothesis. Obviously, whenever the algorithm makes a prediction mistake, at least one hypothesis is removed from .   
Therefore, after making M mistakes we have .  
Since is always nonempty (by the realizability assumption it contains h\*) we have . Rearranging, we obtain the mistake bound.  
Let and let ,  
where ]. Let , , t = 1...d.   
The algorithm might predict for every t ∈ [d]. The number of mistakes made by the algorithm in this case is .   
Hence given condition is not a strict inequality.

**Question II: Stochastic online k-means algorithm**

**Solution:** Given:

take derivative on K-means objective function w.r.t.

Robbins-Monro procedure defines a sequence of successive estimates for   
)  
by replacing with in

Therefore, we get.

Here, is prototype nearest to and the factor of 2 has been absorbed into .

**III: Gaussian mixture model with identical variance**

**Solution:** The mixture of univariate Gaussians with same variance is given as,

In this case the expected maximum of likelihood function becomes w.r.t to of the Gaussian components to zero.

Where, is .

Differentiating this w.r.t and using the below two conditions i.e.

and we get,

for all n. Setting this equal to zero and rearranging we get,

**IV: Biased estimation of Gaussian variance**

**Solution:** We must prove the below condition in order to prove that MLE of variance of a Gaussian distribution is biased.

Where m is the number of i.i.d random variable.

Consider,

and

For i .

Now will be,

From above two equation, we can replace ………………..(1)

Therefore, will become,

Since - and from (1), we get,

**V: Regularized Maximum Likelihood**

**Solution:** The RLM problem is,

Simply add to the training sequence one positive and one negative example, denoted by . Note that the corresponding probabilities are - . Hence minimizing the RLM objective w.r.t original set is same as minimizing ERM w.r.t new set.

Now, MLE is given as -

Given hint is:

We bind RHS inequalities. Note that

We will have two inequalities,

And,

Applying Hoeffding’s inequality hence we get,

Thus, with given confidence parameter 1- delta.

**VI: Bernoulli mixture**

**Solution:** Consider a set of d binary variables , where j= 1…… d which is governed by a Bernoulli distribution with parameter , so that

Where .

Consider a finite mixture of these distributions given by

Where, and

Taking product of the equations given in question, we get.

Now if we marginalize the above equation over z,

Hence the latent variable formulation of a mixture of Bernoulli distribution is equivalent to the product of equation given in question.