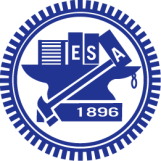
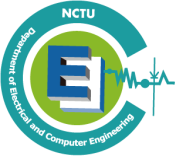
** Machine Learning (Homework #2) **

Due date: 11/27

1. **Information Theory**
2. Please show that the maximum entropy distribution for a continuous variable with three constrains

is a Gaussian distribution.

1. Gaussian distribution is given by

Please derive the corresponding entropy.

1. **Bayesian Inference for the Gaussian**

We develop a Bayesian learning by introducing prior distributions to estimate Gaussian parameters and. Traditionally, batch learning is performed by using the whole training set where high computational complexity is caused. If training data is sufficiently large, it is suitable to use sequential learning (on-line learning) algorithm. Please solve the following question. The file **r2.mat**contains a 1000-point sequence, which is generated by the following multivariate Gaussian distribution with and  **(** is unknown**)**. The sequential learning of the posterior distribution of with the contribution from the final data can be expressed as follows:

1. Please derive the posterior distribution of precision matrix , , in details where is called the *degrees of freedom* of the distribution and is a symmetric matrix. Here, we apply the conjugate prior of whichisa*Wishart* distribution .
2. Please consider the *Wishart* prior and find the MAP solution of (or ) for =10, 100, and 500. () You may also directly use the Matlab command 'wishrnd' to generate many samples of and compare their corresponding to obtain the approximate MAP solution.
3. **Bayesian Inference for the Binomial**

A discrete variable is given with two possible states. Suppose we draw this variable *N* times, the outcomes of the *N* trials are recorded as **O.mat**. Let  denote the numbers of occurrences of two states from the draws. These draws can be represented by a binomial distribution  where denotes the probability or parameter of the first state which satisfies . Please solve the following problems.

1. Please apply the conjugate prior of , whichisaBeta distribution, , derive the posterior distribution , and show the derivation of MAP solution  in details.

(b) **Programming**:

You can use Beta random variable for parameter . Please use the recorded data **O.mat** and plot the prior and posterior distributions from 50 data samples and from the whole data samples. The parameters of the prior distribution are given as .

