Machine Learning (Homework 2)

Due date : 11/30

1 Bayesian Linear Regression (40%)

In this exercise, you will implement an example of Bayesian linear regression and discuss the issues on parameter distribution and predictive distribution.

Dataset:
The file 1_data.mat contains two sequences $\mathbf{x} = \{x_1, x_2, \ldots, x_{100} | 0 \leq x_i \leq 4\}$ and $\mathbf{t} = \{t_1, t_2, \ldots, t_{100}\}$ which represent the input sequence and the corresponding target sequence, respectively.

Basis Function:
Please apply the sigmoidal basis functions $\boldsymbol{\phi} = [\phi_0, \ldots, \phi_{M-1}]^\top$ of the form $\phi_j(x) = \sigma(\frac{x-\mu_j}{s})$ where $\sigma(a)$ is the logistic sigmoid function defined in (3.6). In this exercise, please take the following parameter settings for your basis functions: $M = 7, s = 0.1$ and $\mu_j = \frac{4j}{M}$ with $j = 0, 1, \ldots, (M-1)$. In order to discuss how the amount of training data affects the regression process, please take the data size to be $N = 10, 15, 30$ and 80, for each of the following questions:

1. Please compute the mean vector $\mathbf{m}_N$ and the covariance matrix $\mathbf{S}_N$ for the posterior distribution $p(\mathbf{w}|\mathbf{t}) = \mathcal{N}(\mathbf{w}|\mathbf{m}_N, \mathbf{S}_N)$ with the given prior $p(\mathbf{w}) = \mathcal{N}(\mathbf{w}|\mathbf{m}_0 = 0, \mathbf{S}_0^{-1} = 10^{-6}\mathbf{I})$. The precision of likelihood function $p(\mathbf{t}|\mathbf{w}, \beta)$ or $p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \beta)$ is chosen to be $\beta = 1$.

2. Similar to Fig. 3.9, please generate five curve samples from the parameter posterior distribution.

3. Similar to Fig. 3.8, please plot the predictive distribution of target value t and show the mean curve and the region of variance with one standard deviation on either side of the mean curve.

2 Logistic Regression (60%)

You are given the "seeds" data set (train.csv and test.csv). This data set contains 3 classes. The first 3 dimensions in train.csv are the values of 1-of-K coding for a target vector, the other dimensions are the values of data. In this exercise, as referred in Section 4.3.4, you will implement the Newton-Raphson algorithm to construct a multiclass logistic regression model with the softmax transformation as $p(C_k|\boldsymbol{\phi}) = y_k(\boldsymbol{\phi}) = \exp(a_k)/\sum_j \exp(a_j)$. The error function is formed by using the cross-entropy function as $E(\mathbf{w}) = -\sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk} \ln y_{nk}$. Note: You have to set a stopping criterion $E(\mathbf{w}) < \epsilon$.

1. Set the initial $\mathbf{w}$ to be zero, and show the learning curve of $E(\mathbf{w})$ and the accuracy of classification versus the number of epochs until convergence of training data.

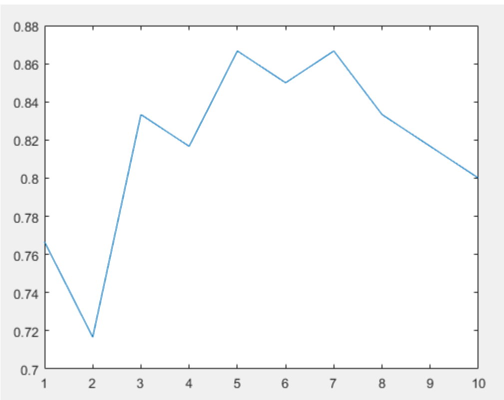
2. Show the classification result of test data.

3. Please plot the distribution (or histogram) of the variable in each dimension of training data and map different colors to each class.
4. Explain that how do you know the model you trained is on the way to global minimum.
5. Please choose a pair of the most contributive variables and plot the samples in training data via 2D graph.
6. Use the variables you choose in (5) and redo (1) and (2).
7. Use the Fisher's linear discriminant (or the linear discriminant analysis) in Section 4.1 to project the data on a two-dimensional (2D) space and plot the training samples in a 2D graph.

3 Nonparametric Methods (Bonus Question 30%)

You are given the "seeds" data set (seeds.csv). This data set contains 3 classes. The last dimension in seeds.csv means the label of class (target values), the other dimensions imply the features of data. In this exercise, you will implement K-nearest-neighbor to construct a multiclass classification model which is referred in Section 2.5 and summarized in the following steps

1. K-Nearest-Neighbor Classifier
   - step 1. There are 70 data samples in each class. You should use the first 50 samples in each class as training data, and the last 20 as test data. You will have 150 training samples from three classes.
   - step 2. You need to preprocess all features by subtracting the mean and normalizing by standard deviation. (formula : $\frac{x-\mu}{\sigma}$)
   - step 3. In inference stage, you compare each test sample with 150 training samples and measure the Euclidean distance between them.
   - step 4. You can use the class with the largest number of occurrences for those K closest training samples to test sample as the prediction of this test sample.
   - step 5. Try different K (from 1 to 10).
   - step 6. Compute the accuracy, i.e. the percentage of correct predictions among 60 test samples).
   - step 7. Plot the figure of accuracy where horizontal axis is K and vertical axis is accuracy.



2. Alternative solution by fixing the distance and determining the K from training data
   - step 1,2,3. Same as above.

- step 4. You can evaluate all training samples and find those samples with the distance to test sample which is smaller than V. You can count those samples and predict the class of this test sample with the largest number of occurrences.
- step 5. Try different V (from 2 to 10).
- step 6. Compute the accuracy.
- step 7. Plot the figure of accuracy where horizontal axis is V and vertical axis is accuracy.