Korea Advanced Institute of Science and Technology
School of Electrical Engineering
EE488 Intro. to Machine Learning, Spring 2018

Issued: Apr. 6, 2018
Due: Apr. 13, 2018

Assignment II-part2

Policy

Group study is encouraged; however, assignment that you hand-in must be of your own work. Anyone suspected of copying others will be penalized. The homework will take considerable amount of time so start as soon as possible..

1. **LDA**: Consider data generated in the following manner: (1) toss a coin with probability of heads equal to p; (2) when the coin toss results in "head" then $y = 1$ otherwise $y = 0$; (3) sample $x|y = 0 \sim N(\mu_0, \Sigma_0)$ and $x|y = 1 \sim N(\mu_1, \Sigma_1)$. Let $\theta = p, \Sigma_1, \Sigma_2, \mu_0, \mu_1$.

   (i) Show that $p(y = 1|\mathbf{x}; \theta) = \frac{1}{1+\exp(f(\theta,\mathbf{x}))}$. Find an explicit form for $f(\theta, \mathbf{x})$.

   (ii) The above represents Linear Discriminant Analysis (LDA) studied in class and it is in similar form as logistic regression. What are the difference and similarity between the two?

   (iii) Consider a two dimensional data $\mathbf{x} \in \mathbb{R}^2$. What would the decision boundary look like for the following cases.

   a. $\mu_0 \neq \mu_1$ but $\Sigma_0 = \Sigma_1 = I$
   b. $\mu_0 \neq \mu_1$ and $\Sigma_0 = I, \Sigma_1 = 4 * I$
   c. $\mu_0 = \mu_1$ and $\Sigma_0 = I, \Sigma_1 = 4 * I$
   d. $\mu_0 = \mu_1 = (0, 0)$ and $\Sigma_0 = \begin{bmatrix} 4 & 1 \\ 1 & 4 \end{bmatrix}, \Sigma_1 = \begin{bmatrix} 4 & -1 \\ -1 & 4 \end{bmatrix}$

2. **Kernel SVM**: In class, Gaussian radial basis function (RBF) Kernel was introduced where $k(\mathbf{x}, \mathbf{z}) = \exp\left(-\frac{1}{2}\|\mathbf{x} - \mathbf{z}\|^2\right) = \phi(\mathbf{x})^T\phi(\mathbf{z})$. Show that nth element of $\phi(x_i)_n = \frac{1}{\sqrt{n!}}\exp\left(-\frac{x_i^2}{2}\right)x_i^n$.

3. **Kernel SVM**: Consider the constrained optimization problem for obtaining soft-SVM

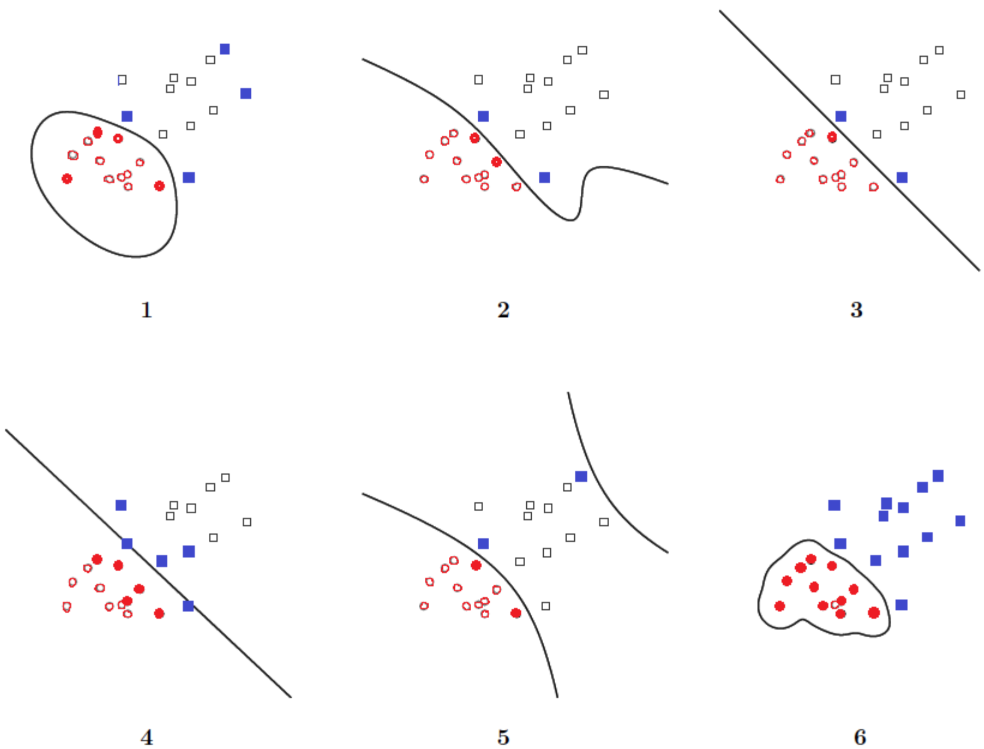
$$\min \frac{1}{2}\|\theta\|^2 + C\sum_{i=1}^{n}\xi_i$$

   such that $y_i(\theta \cdot \mathbf{x_i} + \theta_0) \geq 1 - \xi_i$ and $\xi_i \geq 0$ for $i = 1, \ldots, n$.

   The following figures represents plots of Kernel SVM decision boundaries obtained using different kernels and/or different slack penalties. In the figure, there are two classes of training data with labels $y_i \in \{-1, 1\}$, represented by red circles and blue squares. The filled-in circles and squares represent support vectors. Find from figure $1 - 6$, a figure that presents the following conditions.

   (i) Soft-margin linear SVM with trade-off factor of $C = 0.1$.

   (ii) Soft-margin linear SVM with trade-off factor of $C = 10$.



(iii) Hard-margin kernel machine with $k(\mathbf{x}, \mathbf{z}) = \mathbf{x} \cdot \mathbf{z} + (\mathbf{x} \cdot \mathbf{z})^2$.

(iv) Hard-margin kernel machine with $k(\mathbf{x}, \mathbf{z}) = \exp\left(-\frac{1}{4}\left\|\mathbf{x} - \mathbf{z}\right\|^2\right)$.

(v) Hard-margin kernel machine with $k(\mathbf{x}, \mathbf{z}) = \exp\left(-4\left\|\mathbf{x} - \mathbf{z}\right\|^2\right)$.

4. **Kernel SVM (Matlab Programming Assignment)**: In this problem, we implement a very simple algorithm for solving the following optimization problem of Kernel SVM in the feature spaces $\phi(\mathbf{x})$,

$$\min_{\theta}\left(\frac{\lambda}{2}||\theta||^2 + \frac{1}{m}\sum_{i=1}^{m}\max\{0, 1 - y_i\langle\theta, \phi(\mathbf{x}_i)\rangle\}\right) = \min_{\theta}\left(\frac{\lambda}{2}||\theta||^2 + \mathcal{L}_s(\theta)\right).$$

while only using kernel evaluations where a kernel function is defined as $K(\mathbf{x}_i, \mathbf{x}) = \phi(\mathbf{x}_i)^T\phi(\mathbf{x})$. The basic observation is that the vector $\boldsymbol{\theta}$ maintained by the SGD procedure is always in the linear span of $\{\phi(\boldsymbol{x}_1), \phi(\boldsymbol{x}_2), ..., \phi(\boldsymbol{x}_m)\}$. Therefore, rather than maintaining $\boldsymbol{\theta}^{(t)}$, we can maintain the corresponding coefficients $\boldsymbol{\alpha}$. The algorithm for solving above optimization is given below.

(i) Implement the algorithm in Matlab code.

(ii) Explain line 5 of the algorithm.

Algorithm 1: SGD for Solving Soft-SVM with Kernels

Goal : Solve Equation
Parameter : T
Initialization: $\boldsymbol{\beta}^{(1)} = 0$
for $t = 1, ..., T$ **do**
  Let $\boldsymbol{\alpha}^{(t)} = \frac{1}{\lambda t}\boldsymbol{\beta}^{(t)}$
  Choose i uniformly at random from $[m]$(m is the number of data)
  For all $j \neq i$ set $\beta_j^{(t+1)} = \beta_j^{(t)}$
  if $(y_i \sum_{j=1}^{m} \alpha_j^{(t)} K(\boldsymbol{x}_j, \boldsymbol{x}_i) < 1)$ **then**
    Set $\beta_i^{(t+1)} = \beta_i^{(t)} + y_i$
  else
    Set $\beta^{(t+1)} = \beta^{(t)}$
  end
end
Output : $\bar{\boldsymbol{\theta}} = \sum_{j=1}^{m} \bar{\alpha}_j \phi(\boldsymbol{x}_j)$ where $\bar{\boldsymbol{\alpha}} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\alpha}^{(t)}$

For implementing the RBF kernel svm algorithm, a skeleton code is provided in the "kernel_svm" folder. The main function `main_svm_rbf.m` calls the following 5 functions:
(1) `[x_train,y_train,x_test,y_test] = createDataset('train.csv','test.csv')`,
(2) `[avg_alpha], = train_rbf(x_train,y_train,T)`,
(3) `[acc] = test_rbf(x_test,y_test,x_train,avg_alpha)` and
(4) `plot_svm_rbf(x_train,y_train,x_test,avg_alpha)`
(5) `[K] = svm_kernel(x_i,x_j,kernel)` of radial basis function (RBF) kernels.
(6) `[prediction] = estimate_svm_rbf(x_input,x_train,avg_alpha)` of radial basis function (RBF) kernels.

(i) `[x_train,y_train,x_test,y_test] = createDataset('train.csv','test.csv')` loads both the training and test dataset contained respectively in `train.csv` and `test.csv`.

(ii) `[avg_alpha] = train_rbf(x_train,y_train,T)` takes both the training set {`x_train,y_train`} where `x_train` is the data array while `y_train` is the corresponding label array of training dataset as inputs to the svm and outputs an averaged alpha `avg_alpha`.

(iii) `[acc]=test_rbf(x_test,y_test,x_train, avg_alpha)` takes test set {`x_test,y_test`} the estimated averaged alpha `avg_alpha` and training set {`x_train`} and estimates the classification accuracy `acc`.

(iv) `plot_svm_rbf(x_train,y_train,x_test,y_test,avg_alpha)` takes the training set {`x_train,y_train`} to scatter plots the loaded data and makes the use of the averaged alpha `avg_alpha` to draw a decision boundary on the scatter plot. The positive samples should be represented as red dots while the negative samples should be in blue dots.

(v) `[K] = svm_kernel(x_i, x_j,kernel)` takes the randomly chosen ith sample and $j \neq i$th sample and kernel type, {`x_i, x_j, kernel`} and outputs a feature map `kernel` is defined by {`'RBF'`}.

(vi) `[y] = estimate_kernel(x_grid, x_train,avg_alpha)` takes the `x_grid` data samples, `x_train, avg_alpha` and outputs prediction scores `y`. In the skeleton code, the additional information on `x_grid` given.

Submit Instructions for Programming Assignment

- Please submit in .zip file to KLMS named **ee488_assignment2_student#.zip**, for example, "ee488_assignment2_20181234.zip".

- This file should contain the following folder - **kernel_SVM** and each folder contains document file for the result with analysis.

- In matlab code, the comment explaining your code **must be** included, or you will not get a full grade even if your code works fine. Please also include all the files that are required to run the code in the zip file. Do not change the name of the folder and comments should be written in English. Additionally submitting unexecutable code will receive no points.