Machine Learning, Spring 2018
Homework 5

Due on 23:59 May 16, 2018
Compress all your materials in one file, and send to $cs282\_01@163.com$
with subject "Chinese name+student number+HW5" (In this format please!)

Understanding VC dimension (20 points)

In this part, you need to complete some mathematical proofs about VC dimension. Suppose the hypothesis set

$$\mathcal{H} = \{f(x, \alpha) = \text{sign}\ (\sin(\alpha x))|, \alpha \in \mathbb{R}\}$$

where x and f are feature and label, respectively.

- Show that $\mathcal{H}$ cannot shatter the points $x_1 = 1, x_2 = 2, x_3 = 3, x_4 = 4$. (8 points)

  (Key: Mathematically, you need to show that there exists y_1, y_2, y_3, y_4, for any $\alpha \in \mathbb{R}$, $f(x_i) \neq y_i, i = 1, 2, 3, 4$, for example, $+1, +1, -1, +1$)

- Show that the VC dimension of $\mathcal{H}$ is ∞. (Note the difference between it and the first question) (12 points)

  (Key: Mathematically, you have to prove that for any label sets $y_1, \cdots, y_m, m \in \mathbb{N}$, there exists $\alpha \in \mathbb{R}$ and $x_i, i = 1, 2, \cdots, m$ such that $f(x; \alpha)$ can generate this set of labels. Consider the points $x_i = 10^{-i}$...)

Understanding Lasso (30 points)

Consider the following generalized Lasso problem

$$\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{Ax} - \boldsymbol{b}\|_2^2 + \lambda \|\boldsymbol{Fx}\|_1, \tag{1}$$

where $\boldsymbol{A}$ is the under-determined sensing matrix, $\boldsymbol{F}$ is the transformed matrix. In particular, it can be reduced to Lasso problem if $\boldsymbol{F} = \boldsymbol{I}$. The above problem is equivalent to the following formulation

$$\begin{aligned} &\min_{\boldsymbol{x},\boldsymbol{z}} \frac{1}{2} \|\boldsymbol{Ax} - \boldsymbol{b}\|_2^2 + \lambda \|\boldsymbol{z}\|_1 \\ &\text{s.t. } \boldsymbol{Fx} = \boldsymbol{z}, \end{aligned} \tag{2}$$

and one can employ augmented Lagrangian multiplier method to solve it. Specifically, the augmented Lagrangian is

$$\mathcal{L} = \frac{1}{2} \|\boldsymbol{Ax} - \boldsymbol{b}\|_2^2 + \lambda \|\boldsymbol{z}\|_1 + < \boldsymbol{y}, \boldsymbol{Fx} - \boldsymbol{z} > + \frac{1}{2}\rho \|\boldsymbol{Fx} - \boldsymbol{z}\|_2^2,$$

which yields the ADMM algorithm, see Algorithm 1. The soft-thresholding operator $\mathbb{S}_{\frac{\lambda}{\rho}}$ is defined as

$$\mathbb{S}_{\frac{\lambda}{\rho}}(x_i) = \begin{cases} x_i - \frac{\lambda}{\rho},\ x_i \geq \frac{\lambda}{\rho} \\ 0,\ |x_i| < \frac{\lambda}{\rho} \\ x_i + \frac{\lambda}{\rho},\ x_i \leq -\frac{\lambda}{\rho}. \end{cases} \tag{3}$$

Algorithm 1 ALM for generalized Lasso problem

Input: $\boldsymbol{A}, \boldsymbol{F}, \boldsymbol{b}, \lambda, \mu$ (for augmented Lagrange multiplier)

Initialized $\rho, \boldsymbol{x}_0, \boldsymbol{z}_0$, and $\boldsymbol{y}_0$, $k = 0$

while not converged **do**

  $\boldsymbol{x}^{(k+1)} = \textit{update } \boldsymbol{x}?$

  $\boldsymbol{z}^{(k+1)} = \textit{update } \boldsymbol{z}?$ (you may want to use $\mathbb{S}_{\frac{\lambda}{\rho}}$ element-wise)

  $\boldsymbol{y}^{(k+1)} = \boldsymbol{y}^{(k)} + \rho(\boldsymbol{F}\boldsymbol{x}^{(k+1)} - \boldsymbol{z}^{(k+1)})$

  $\rho = \rho\mu$

  $k = k + 1$

end while

Output: $\boldsymbol{x}^* = \boldsymbol{x}^{(k)}$

- Derive the steps of update $\boldsymbol{x}$ and $\boldsymbol{z}$ in Algorithm 1. (10 points)
- Complete the function `glasso.m`, and pass the test using `testglasso.m` (15 points)
- Run `demo.m` and report your MSE and PSNR. (5 points)

Dual Formulation of the SVM (25 points)

Compared with the SVM formulation in Lecture 15 and 16, its dual problem will be much eaiser, since the original problem has so many constraints.

- Give the dual formulation of the SVM and what the KKT condition is for primal SVM. We need you show the induction of the procedure.
- There is a efficient for dual SVM problem, called Sequential Minimal Optimization. You can find many materials for this algorithm from website, it make use of the KKT conditios to solve the dual quadratic problem.
  - Give us an abstract of its principle and the pseudocode of the SMO algorithm for the dual SVM problem.
  - *a3a* is a data for the binary classification, show us your accuracy of the test data, like Fig 1. You need download the data *a3a* from https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/

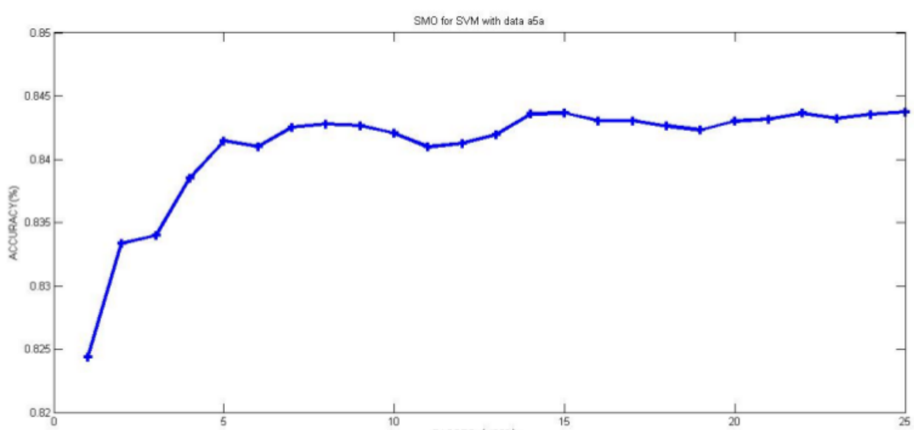

Figure 1: Accuracy demo.

Kernel function $(25$ points)

Suppose we are given the following positively ("$+1$") labeled data points $\mathbb{R}^2$:

$$\left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} -1 \\ -1 \end{pmatrix} \right\},$$

and the following negatively labeled ("-1") data points in $\mathbb{R}^2$ (see Figure 2):

$$\left\{ \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \quad \begin{pmatrix} 2 \\ -2 \end{pmatrix}, \quad \begin{pmatrix} -2 \\ 2 \end{pmatrix}, \quad \begin{pmatrix} -2 \\ -2 \end{pmatrix} \right\}.$$

Question:

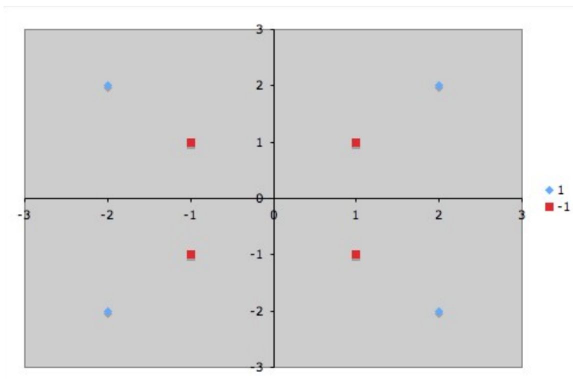

Figure 2: Blue diamonds are positive examples and red squares are negative examples.

1. Find a kernel function to map the data into a $\mathbb{R}^3$ feature space and make the new data linearly separable. (Show your kernel function please!) (5points)
2. Use SVM classifier to seperate the data. Show the SVM problem in your report. (10points) Solve this SVM problem, write down the expression of the final separating hyperplane, and plot the data and the separating hyperplane in a figure. (10points) (You can solve the SVM problem by apllying a convex problem solver.)