**CSE 847 (Spring 2016): Machine Learning— Homework 3**

Instructor: Jiayu Zhou

Due on Tuesday, March 15, before class.

# 1 Linear Algebra III

1. Let $A \in \mathbb{R}^{m \times n}$ be a matrix of rank $r$. Prove that $\|A\|_F \leq \sqrt{r}\|A\|_2$. Here $\|A\|_F$ denotes the Forbenius norm of $A$.
2. Let $A \in \mathbb{R}^{m \times n}$ be a matrix of rank $n$. Prove that $\|A(A^TA)^{-1}A^T\|_2 = 1$.
3. Let $A$ and $B$ be two positive semi-definite matrices in $\mathbb{R}^{n \times n}$. Prove or disprove:
   (a) $A + B$ is positive semi-definite
   (b) $AB$ is positive semi-definite
   (c) $B^T$ is positive semi-definite

# 2 Linear Classification

Questions in the textbook Pattern Recognition and Machine Learning:

1. Page 220, Question 4.1
2. Page 221, Question 4.5
3. Page 221, Question 4.6
4. Page 222, Question 4.15

# 3 Linear Regression: Experiment

In this part of homework you will explore the ridge regression and the effects of $\ell_2$-norm regularization. You are to implement a MATLAB solver for ridge regression:

$$\min_w \frac{1}{2}\|Xw - y\|_2^2 + \frac{\lambda}{2}\|w\|_2^2.$$

You are not allowed to use the integrated ridge regression in MATLAB. You will use your solver to investigate the effects of the regularization on the DIABETES [1] dataset, and study the cross validation procedure.

1. Implement the ridge regression solver.
2. Train regression models on the DIABETES dataset using the training data (x_train, y_train variables in the data file). Vary the $\lambda$ from $1e-5, 1e-4, 1e-3, 1e-2, 1e-1, 1, 10$ (In Matlab $1e-1$ means 0.1). Compute training error (predict y_train given X_train), testing error (predict y_test given X_test) for each $\lambda$. The error is measured by mean squared error (MSE):

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2,$$

[1] https://github.com/jiayuzhou/CSE847-2016Spring/blob/master/homework/diabetes.mat?raw=true

where $N$ is the number of samples on which the error is computed, $y_i$ is ground truth, and $\hat{y}_i$ is the prediction from data points given model $w$.

3. Perform 5-fold cross validation on the training data to estimate the best $\lambda$ from training data.

In the homework, attach a brief report. In the report you need to discuss your findings in the experiment, include a plot showing how training/testing error changes when you vary the parameter $\lambda$ (use log scale on $\lambda$). In the same plot, show the best $\lambda$ obtained from your 5-fold cross validation procedure. Submit the MATLAB code (do add some comments in your code for others to understand your code) to a public repository under your Github account (the same account of your project) and include the link in the report.