CSE 847 (Spring 2021): Machine Learning— Homework 3
Instructor: Jiayu Zhou
Due on Wednesday, Mar 17 11:59 PM Eastern Time.

1 Linear Algebra III

1. (10 points) Let $A \in \mathbb{R}^{m\times n}$ be a matrix of rank n. Prove that $\|A(A^TA)^{-1}A^T\|_2 = 1$.
2. (10 points) 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. (10 points) Page 220, Question 4.1
2. (10 points) Page 221, Question 4.5
3. (10 points) Page 221, Question 4.6
4. (10 points) Page 222, Question 4.15

3 Linear Regression: Experiment

(40 points) In this part of homework you will explore the ridge regression and the effects of ℓ_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 λ 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 λ. The error is measured by mean squared error (MSE):

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

   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.

[1] https://github.com/jiayuzhou/CSE847/blob/master/data/diabetes.mat?raw=true

3. Perform 5-fold cross validation on the training data to estimate the best λ 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 λ (use log scale on λ). In the same plot, show the best λ 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 the D2L along with your report.

Response: For the regression problem, we can find the expression of the weights as:

$$w = (\Phi^T \Phi + \lambda \mathbb{I})^{-1} \Phi^T t$$

where Φ is the design matrix and t is the target vector. The weights are computed from the training matrix and then applied over test data to compute the predictions. Based on the predictions, every such model provides a MSE which can be used as the quality indicator of the model.

The model has one hyperparamter which is λ. In order to get one optimal value of λ, k-fold cross-validation is performed. After getting an optimal value of λ, that is used to build the model which is finally applied over the test data. The entire result of this experimentation is tabulated in Table 3.

Different types of normalization techniques are also used on the data to measuer the change in errors based on different types of normalizations. For this purpose 5 types of normalization techniques are used apart from the non-normalized scenario. These normalization techniques are available in *MATLAB* under the names: 'range', 'scale', 'zscore', 'norm', 'center'. Each of these normalization techniques use different methods to scale the data.

	Error Type	**Normalization Type**					
		None	**Range**	**Scale**	**Z-score**	**Norm**	**Center**
λ=1e-6	Train Error	19560.40391	0.021933	3.366411	0.385461	0.002891	2189.03486
	Test Error	93835.00672	3.589247	1005.224421	4.338947	0.056803	3558.254969
λ=1e-5	Train Error	19695.35482	0.021947	3.380637	0.385516	0.002906	2194.171893
	Test Error	76911.07209	1.621412	462.882938	2.244155	0.032902	3514.411567
λ=1e-4	Train Error	20060.11614	0.022058	3.431649	0.385708	0.002956	2209.43153
	Test Error	57531.76027	0.123865	80.669288	0.808274	0.013571	3483.224263
λ=1e-3	Train Error	20384.66425	0.022171	3.451689	0.385887	0.003003	2227.638985
	Test Error	49060.59804	0.107295	40.584322	0.649208	0.0083	3285.303421
λ=1e-2	Train Error	20918.86518	0.022419	3.493135	0.387634	0.003069	2306.837688
	Test Error	44405.1394	0.055242	18.906226	0.644382	0.007227	2951.255302
λ=1e-1	Train Error	22075.71567	0.023389	3.556196	0.390146	0.003194	2511.476379
	Test Error	36994.69673	0.035062	9.140952	0.636367	0.006359	2852.421846
λ=1	Train Error	24730.73933	0.025154	3.620571	0.39539	0.003527	3288.941082
	Test Error	30517.77394	0.029653	7.306174	0.543775	0.005187	3510.851314
λ=10	Train Error	27165.12613	0.03021	3.728976	0.415413	0.003947	4917.706446
	Test Error	30057.3035	0.033097	6.58842	0.470815	0.004905	5390.676364
λ=100	Train Error	27827.97731	0.045432	4.009522	0.465563	0.004107	5556.667563
	Test Error	30414.29643	0.048561	5.430243	0.474178	0.004982	6136.625862
Best λ according to CV-MSE		10	1	30	30	10	0.1
MSE for Non-Regularized version		106775.3616	3.972377	1132.597799	4.838317	1.160637	3630.528184
MSE for Ridge Regularized version		30057.3035	0.029653	6.081624	0.461673	0.004905	2852.421846
MSE for MATLAB integrated Ridge Regularized version		30145.62543	0.094501	4.25334	0.461673	0.004928	5911.960252