

There are 8 problems in this set. You must complete 6 (doing more will get no credit - work on your project!) 3 of the problems (you choose) are due on September 12, and the rest of the problems you complete are due on September 19. Feel free to work with other students, but make sure you write up the homework and code on your own (no copying homework *or* code; no pair programming). Feel free to ask students or instructors for help debugging code or whatever else, though. When implementing algorithms you may not use any library (such as sklearn) that already implements the algorithms but you may use any other library for data cleaning and numeric purposes (numpy or pandas). Use common sense. Problems are in no specific order.

**1 (regression).** Download the data at [https://math189r.github.io/hw/data/online\\_news\\_popularity/online\\_news\\_popularity.csv](https://math189r.github.io/hw/data/online_news_popularity/online_news_popularity.csv) and the info file at [https://math189r.github.io/hw/data/online\\_news\\_popularity/online\\_news\\_popularity.txt](https://math189r.github.io/hw/data/online_news_popularity/online_news_popularity.txt). Read the info file. Split the csv file into a training and test set with the first two thirds of the data in the training set and the rest for testing. Of the testing data, split the first half into a ‘validation set’ (used to optimize hyperparameters while leaving your testing data pristine) and the remaining half as your test set. We will use this data for the remainder of the problem. The goal of this data is to predict the **log** number of shares a news article will have given the other features.

(a) (**math**) Find a closed form solution  $\mathbf{x}^*$  to the ridge regression problem:

$$\text{minimize: } \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \|\Gamma\mathbf{x}\|_2^2.$$

(b) (**implementation**) Attempt to predict the log shares using ridge regression from the previous problem solution. Make sure you include a bias term and *don't regularize the bias term*. Find the optimal regularization parameter  $\lambda$  from the validation set. Plot both  $\lambda$  versus the validation RMSE (you should have tried at least 150 parameter settings randomly chosen between 0.0 and 150.0 because the dataset is small) and  $\lambda$  versus  $\|\boldsymbol{\theta}^*\|_2$  where  $\boldsymbol{\theta}$  is your weight vector. What is the final RMSE on the test set with the optimal  $\lambda^*$ ?

(c) (**math**) Consider regularized linear regression where we pull the bias term out of the feature vectors. That is, instead of computing  $\hat{\mathbf{y}} = \boldsymbol{\theta}^\top \mathbf{x}$  with  $\mathbf{x}_0 = 1$ , we compute  $\hat{\mathbf{y}} = \boldsymbol{\theta}^\top \mathbf{x} + b$ . This corresponds to solving the optimization problem

$$\text{minimize: } \|\mathbf{A}\mathbf{x} + b\mathbf{1} - \mathbf{y}\|_2^2 + \|\Gamma\mathbf{x}\|_2^2.$$

Solve for the optimal  $\mathbf{x}^*$  explicitly. Use this close form to compute the bias term for the previous problem (with the same regularization strategy). Make sure it is the same.

(d) (**implementation**) Use the solution