

Homework # 4

Due Thursday, Th Apr 7th 12:30pm.

NAME: _____

Signature: _____

STD. NUM: _____

General guidelines for homeworks:

You are encouraged to discuss the problems with others in the class, but all write-ups are to be done on your own.

Homework grades will be based not only on getting the “correct answer,” but also on good writing style and clear presentation of your solution. It is your responsibility to make sure that the graders can easily follow your line of reasoning.

Try every problem. Even if you can't solve the problem, you will receive partial credit for explaining why you got stuck on a promising line of attack. More importantly, you will get valuable feedback that will help you learn the material.

Please acknowledge the people with whom you discussed the problems and what sources you used to help you solve the problem (e.g. books from the library). This won't affect your grade but is important as academic honesty.

When dealing with python exercises, please attach a printout with all your code and show your results clearly.

The purpose of these exercises will be to apply ML techniques learned during the course to competitions in Kaggle. A working algorithm in each question will guarantee full marks. Winning a competition will guarantee full marks for the homeworks.

Try **two** of the following:

1 UW STAT331 Linear Models Contest

Go to <http://inclass.kaggle.com/stat331w11> get the data and try the Bayesian linear models, with priors on the variance and the parameters. That is, the predictive distribution should be a t-distribution as Kevin's book. Kevin also has a chapter where he describes *Lasso*, which corresponds to an L_1 regulariser (we covered this when we talked about convex optimization). Try this out. Beyond this, Kevin has a whole chapter on L_1 regularization. Peter Buhlman also has lots of nice papers on the topic. What RMSE did you get?

2 UMICH SI650 - Sentiment Classification

Go to <http://inclass.kaggle.com/si650winter11> get the data and try the classifiers we considered in class: logistic regression, GP classifiers, CRFs, k-nearest neighbors or anything you could come up with. Sentiment classification is what finance people use to predict whether the feeling in the media is that, say gold, will go up. Politicians like knowing what people think of them too. So do marketers.

3 UCL Applied Machine Learning 2 - Film recommendation problem

Go to <http://inclass.kaggle.com/UCLAML2p1> and good luck.