Name	HW6
Description	The final answers are not regarded as evaluation criteria. In other words, two students might get completely different values in one question, but still get full credits each. This is even true about the yes/no or A/B questions, meaning that one student might pick A as the answer and give acceptable description and get full credit while the other student might pick completely the opposite answer, and give sound reasons and yet again get full credit. Therefore, only the final results are not determinant of the student's grade.

Rubric Detail

	Levels of Achievement		
Criteria	**Novice**	**Competent**	**Proficient**
Q1- Building a straightforward linear regression **Weight 10.00%**	**0 %** Not Done.	**80 %** Incomplete/incorrect Plot/R-Square value reports (i.e. two values each). Results/plots exactly similar to reference values/plots are not expected from students, but if the results are wildly different from the reference data that they suggest mis-training of the model, then the answer will be marked as competent.	**100 %** Reporting the R-Squared values, and plotting the fitted models for each label (i.e. latitude and longtitude).

	Levels of Achievement		
Criteria	**Novice**	**Competent**	**Proficient**
Q1-Performing Box-Cox Transformation **Weight 10.00%**	**0 %** Not Done.	**80 %** Performing Box-Cox transformation and plotting the log likelihood versus lambda for just one label (i.e. either latitude or longtitude but not both). Also, not accounting for negative data values for box-cox transformation can make the answer marked as competent.	**100 %** Solving the negative data value somewhat (Easiest way is to convert angle values into positive ones). Performing Box-Cox transformation and plotting the log likelihood versus lambda for both labels (i.e. latitude and longtitude).
Q1- Performing valid model selection test for choosing the best box-cox **Weight 0.00%**	**0 %** Not Done.	**80 %** Using unfit model selection criteria such as MSE or R-Squared.	**100 %** The right way of selection between the best lambda model and the original model is to look at the residuals vs fitted values plot, and make a decision based on the shape of the plot. The value of best lambda itself could be considered when making decision. The final picked model could be the original, or the transformed depending on the explanation.
Q1- Choosing best box-cox lambda value **Weight 10.00%**	**0 %** Not Done.	**0 %** Not Done.	**100 %** Picking a lambda based on MSE or R-Squared or likelihood or any other acceptable criteria.

	Levels of Achievement		
Criteria	**Novice**	**Competent**	**Proficient**
Q1-L2 Regularization **Weight 10.00%**	**0 %** Not Done.	**80 %** Missing one of the necessary parts of the answer (See the proficient column).	**100 %** (A) Using a kind of CV-error vs lambda(or log lambda) plot or a set of Cross-Validation accuracy values for picking the best regularization lambda value (B) Reporting the regularization coefficient that produces the minimum error (both lambda.1se and lambda.min are acceptable if reported) (C) Explaining correctly whether the regularized regression better than the unregularized regression. *** The whole process should be done two times, once for latitude and once for longtitude. 6 Values should be reported at least: (CV Error, unregularized, latitude), (CV Error, regularized, latitude), (CV Error, unregularized, longtitude), (CV Error, regularized, longtitude), (Best Regularization Coeff, lattitude), (Best Regularization Coeff, longtitude)
Q1-L1 Regularization **Weight 15.00%**	**0 %** Not Done.	**80 %** Missing one of the necessary parts of the answer (See the proficient column).	**100 %** (A) Using a kind of cv error vs lambda plot or a set of Cross-Validation accuracy values for picking the best regularization lambda value (B) Reporting the regularization coefficient that produces the minimum error (both lambda.1se and lambda.min are acceptable if reported) (C) Reporting the number of parameters(i.e. variables) after lasso 4) Explaining correctly whether the regularized regression is better than the unregularized regression. (D) The whole process should be done two times, once for latitude and once for longtitude. 10 Values should be reported: (CV Err, unregularized, latitude), (CV Err, regularized, latitude), (CV Err, unregularized, longtitude), (CV Err, regularized, longtitude), (Best Regularization Coef, lattitude), (Best Regularization Coef, longtitude), (# of Parameters, Unregularized, latitude), (# of Parameters, regularized, longtitude), (# of Parameters, Unregularized, longtitude), (# of Parameters, regularized, latitude). A couple sentences for explanation.

	Levels of Achievement		
Criteria	**Novice**	**Competent**	**Proficient**
Q1-Elastic Net Regularization **Weight 20.00%**	**0 %** Not Done.	**80 %** Missing one of the necessary parts of the answer (See the proficient column).	**100 %** (A) Using a kind of cv error vs lambda plot or a set of Cross-Validation accuracy values for picking the best regularization lambda value (B) Reporting the regularization coefficient that produces the minimum error (both lambda.1se and lambda.min and anything between them, are acceptable if reported) (C) Reporting the number of parameters(i.e. variables) after lasso 4) Explaining correctly whether the regularized regression is better than the unregularized regression. (D) The whole process should be done two times, once for latitude and once for longtitude. 8 CV-Error values should be reported ((3 alphas and unregularized) * (two variables)). 6 optimal Values of regularization constant ((3 alphas) * (two variables)). 6 Number of parameters after regularization ((3 alphas) * (two label variables)). A couple sentences of interpretation for these numbers should be present as the reason for deciding to pick regularized or unregularized as the best model.
Q2- Performing Logistic Regression Generally **Weight 5.00%**	**0 %** Not Done.	**80 %** Implementation errors	**100 %** Correctly implemented
Q2- Reporting the optimal values **Weight 10.00%**	**0 %** Not Done.	**80 %** Incompletely done	**100 %** You should report any results pertaining to your choice of regularization (e.g. the plots from glmnet), in addition to the accuracy of your best classifier. Even if you didn't end up using regularization, you should still show evidence that you tried it. Some things you could include are the plots from glmnet, a list of lambda and alpha values that you tried along with their corresponding accuracies/errors, etc.
Q2- Analysis about the best model **Weight 10.00%**	**0 %** Not Done	**80 %** Partially acceptable analysis and reasoning.	**100 %** Picking the right model according to the statistics and most importantly the CV error rate.

View Associated Items

Print Close Window