**CSCE 623: Machine Learning**

**Spring 2019**

**HW4 GRADING WORKSHEET**

Due Tuesday, 14 May at 2359

Submit via Canvas **(**This homework assignment is worth 5 points toward your final grade**)**

|  |  |
| --- | --- |
| Course points earned | 4.95 |

**Student Name: NEWLIN**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Step | Step Evaluation Criteria | Max | Student performance | Studnt Score |
| 1 | Explore data (plot), eyeball & state good features for linear regression; Discuss data distribution & make prediction about useful column features | 1 | Achieved | 1 |
| 2 | Build bestSubset(X\_nonTest,y\_nonTest, k) to compute *MSE* on training set | 2 | Achieved | 2 |
| 3 | Run best subset for all possible model sizes 1 to 6 and produce table of feature values. Discuss interesting outcomes & changes in features as model grows | 1 | Achieved | 1 |
| 4 | Plot val set MSE of each of the 6 best models. Annotate best model & k | 2 | Achieved | 2 |
| 5 | Report val set MSE on model with best k features Discuss findings and compare to to features explored in step 1 | 2 | Achieved | 2 |
| 6 | Code forward-stepwise function for Linear Regression to compute *R*2 on training set forwardStepwiseSubset(X\_nonTest,y\_nonTest, k) | 2 | Achieved | 2 |
| 7 | Run forward stepwise & print out table of stepwise best features of sizes 1-to-6. Discuss how stepwise features change compared to best-selected features in step 5 | 2 | Achieved | 2 |
| 8 | Update plot w/line to the previous plot to include val set class accuracy of each of the p best models. Annotate best performing model in plot & report MSE on the val set for this model | 2 | Achieved | 2 |
| 9 | Describe changes in model MSE for forward stepwise as k grows 1 to 6. Discuss findings and compare to evidence in step 1. | 2 | Partial: if the models are using the same features, then you should explain why they are getting different MSEs. Perhaps this is because your CV is using different random folds? | 1.5 |
| 10 | Discuss outcomes in terms of tradespace (accuracy & computational complexity). Discuss independence/interdependence of the features. | 2 | Achieved | 2 |
| 11 | Write a function LASSOSubset(X\_nonTest,y\_nonTest, k) with 5-fold crossval to determine average cross-val MSE for the set of k best features | 2 | Achieved | 2 |
| 12 | Run LASSO on k=1..6 and present the outputs of the search (e.g. table) | 2 | Achieved | 2 |
| 13 | Add the LASSO line to the plot of performance; annotate best LASSO performance model | 2 | Achieved | 2 |
| 14 | Describe the change as LASSO k grows from one to 6 and discuss with respect to eyeball features | 2 | Achieved | 2 |
| 15 | Customer Questions – recommended input features for model size 1-6;  Best model and feature size and performance | 4 | Achieved | 4 |
| Holistic Grade | Grade = 0 if some recommendations are missing or all recommendations are provided, but at least some of the recommendations provide no supporting evidence.  Grade = 10 if all recommendations are present and provide evidence, but at least some evidence is confusing, misleading, or doesn’t support the conclusion.  Grade = 20 if all recommendations are present and convincing evidence is provided for each conclusion. | | | 20 |
| TOTAL |  | 50 |  | 49.5 |