​**Here is the rubric I will be using to grade the MiniLab.  You must have these exact sections in your project, if you want me to grade it!  The MiniLab is due Sunday, 2/16/2020.**

|  |  |  |
| --- | --- | --- |
| **Category** | **Available** | **Requirements** |
| Total Points | 100 |  |
|  |  |  |
| Create Models | 50 | Create a logistic regression model and a support vector machine model for the classification task involved with your dataset. Assess how well each model performs (use 80/20 training/testing split for your data). Adjust parameters of the models to make them more accurate. If your dataset size requires the use of stochastic gradient descent, then linear kernel only is fine to use. That is, the SGDClassifier is fine to use for optimizing logistic regression and linear support vector machines. For many problems, SGD will be required in order to train the SVM model in a reasonable timeframe. |
| Model Advantages | 10 | Discuss the advantages of each model for each classification task. Does one type of model offer superior performance over another in terms of prediction accuracy? In terms of training time or efficiency? Explain in detail. |
| Interpret Feature Importance | 30 | Use the weights from logistic regression to interpret the importance of different features for the classification task. Explain your interpretation in detail. Why do you think some variables are more important? |
| Interpret Support Vectors | 10 | Look at the chosen support vectors for the classification task. Do these provide any insight into the data? Explain. If you used stochastic gradient descent (and therefore did not explicitly solve for support vectors), try subsampling your data to train the SVC model— then analyze the support vectors from the subsampled dataset. |