	Project Rubric	Status	Comment	
	I Machine Learning Question: 20 pts			
	A. Is the background context for the question stated clearly (with references)?	Check		
	B. Is the hypothesis/problem stated clearly ("The What")	No	some what/ come back later	Ask Professor
	C. Is it clear why the problems are important? Is it clear why anyone would care? ("The Why")	Check		
	D. Is it clear why the data were chosen should be able to answer the question being asked?	Check		
	E. How new, non-obvious, significant are your problems? Do you go beyond checking the easy and obvious?	Check	Can't do anything about it	
	II Data Cleaning/Checking/Data Exploration: 20pts			
	A. Did you perform a thorough EDA (points below included)?	Check	Need to add feature selection portion	HuiHuang
	B. Did you check for outliers?	Check		
	C. Did you check the units of all data points to make sure they are in the right range?	Check		
	D. Did you identify the missing data code?	Check		
	E. Did you reformat the data properly with each instance/observation in a row, and each variable in a column?	Check		
	F. Did you keep track of all parameters and units?	Check		
	G. Do you have a specific code for reformating the data that does not require information not documented (eg. magic numbers)?	Check	Not sure what it mean	
	H. Did you plot univariate and multivariate summaries of the data including histograms, density plots, boxplots?	Check		
	I. Did you consider correlations between variables (scatterplots)?	Check	No scatterplot needed	
	J. Did you consider plot the data on the right scale? For example, on a log scale?	Check		
	K. Did you make sure that your target variables were not contaminating your input variables?	Check		
	L. If you had to make synthetic data was it a useful representation of the problem you were trying to solve?	Check	No need for synthetic data	
	III. Transformation, Feature Selection, and Modeling: 30pts			
	A. Did you transform, normalize, filter the data appropriately to solve your problem? Did you divide by max-min, or the sum, root-square-sum, or did you z-score the data? Did you justify what you did?	Check	Only need to normalize image	
	B. Did you justify normalization or lack of checking which works better as part of your hyper-parameters?	Check	Done	
	C. Did you explore univariate and multivariate feature selection? (if not why not)	Check		
	D. Did you try dimension reduction and which methods did you try? (if not why not)	Check	Done	
	E. Did you include 1-2 simple models, for example with classification LDA, Logistic Regression or KNN?	Check		
	F. Did you pick an appropriate set of models to solve the problem? Did you justify why these models and not others?	Check	Need to explain in report	Gong
	G. Did you try at least 4 models including one Neural Network Model using Tensor-Flow or Pytorch?	Check		
	H. Did you exercise the data science models/problems we described in the lectures showing what was presented?	Check	What does this even mean?	
	I. Are you using appropriate hyper-parameters? For example, if you are using a KNN regression are you investigating the choice of K and whether you use uniform or distance weighting? If you are using K-means do you explain why K? If you are using PCA do you explore how many dimensions such as by looking at the eigenvalues?	Check	Need to explain in report	Gong
	IV. Metrics, Validation and Evaluation 20pts			
	A. Are you using an appropriate choice of metrics? Are they well justified? If you are doing classification do you show a ROC curve? If you are doing regression are you justifying the metric least squares vs. mean absolute error? Do you show both?	Check	Need to explain in report	Gong
	B. Do you validate your choices of hyperparameters? For example, if you use KNN or K-means do you use cross-validation to optimize your choice of parameters?	Check	Need to tuning for CNN	Gong
	C. Did you make sure your training and validation process never used the training data?	Check		
	D. Do you estimate the uncertainty in your estimates using cross-validation?	No	What does it mean?	Gong
	E. Can you say how much you are overfitting?	Check		
	V. Visualization 10pts			
	A. Do you provide visualization summaries for all your data and features?	No	Need to add to report	HuiHuang

	B. Do you use the correct visualization type, eg. bar graphs for categorical data, scatter plots for numerical data, etc?	No	Need to add to report	HuiHuang
	C. Are your axes properly labeled?	No	Need to add to report	HuiHuang
	D. Do you use color properly?	No	Need to add to report	HuiHuang
	E. Do you use opacity and dot size so that scatterplots with lots of data points are not just a mass of interpretable dots?	Check	no scatterplot needed	
	F. Do you write captions explaining what a reader should conclude from each figure (not just saying what it is but what it tells you)?	No	Need to add to report	HuiHuang/Gong
	VI. Code 20pts			
	A. Is the code provided can reproduce the entire work?	No	Need to aggreate the code	Gong
	B. Is the data included or at least linked (externally) with instructions on how to download it?	Check		
	C. Do you factor repeated operations into functions to avoid repetitively and error-prone copy-paste?	No		
	E. Do you use docstrings and numpy documentation style:	No	Need to do	Gong/HuiHuang
	https://github.com/numpy/numpy/blob/master/doc/HOWTO_DOCUMENT.rst.txt			
	to make your code clear and readable?			
	F. Do you use markdown cells to explain every step of your code similar to Homeworks and some example notebooks?	No	Need to do	Gong/HuiHuang
	G. Does the code demonstrate considerable work given the number of people on the project?	Check		
	VII. Presentation 30pts			
	A. Do you tell a coherent story with a beginning, middle, and end?	Check		
	B. Do you introduce why the problem is important?	Check		
	C. Do you explain in the first couple of slides what you accomplished on solving the problem?	Check		
	D. Are you careful not to have slides filled with text (keep in notes)?	Check		
	E. Is data and evaluations presented as clear figures (mostly)?	Check		
	F. Do you make sure to say what is "interesting" or should be learned from each figure?	Check		
	G. Do you stay within your time limits 15 min?	Check		
	H. Do you avoid useless padding slides of no relevance?	Check		
	VIII. Report 30pts			
	The structure should follow this template above. Parts are listed in that section.	In progress		