	Ideal Score	Shen Yang Team: YS Score	Instructor Notes
Total Points	10	8.15	Score Points:08.15
Total Percentage	100	81.5	
Exceptional Work, (**required for 7000 level students**): Implement two more phi functions: ReLU and SiLU (also called Swish). Compare their performance to the linear and sigmoid phi functions.	10	10	Relu and Silu implemented and analyzed.
Explain the task and **what business-case or use-case** it is designed to solve (or designed to investigate). **Detail exactly what the task is and what parties would be interested in the results.** How well would your prediction algorithm need to perform to be considered useful by interested parties?	5	5	Excellent.
Define and prepare your class variables. Use proper variable representations (int, float, one-hot, etc.). **Use pre-processing methods (as needed) for dimensionality reduction, scaling, etc.** Remove variables that are not needed/useful for the analysis. **Describe the final dataset that is used for classification/regression** (include a description of any newly formed variables you created).	10	7	There are a number of codes here that should be one hot encoded, not scaled by a standard scaler.
Choose and explain what metric(s) you will use to evaluate your algorithm's generalization performance. You should give a detailed argument for why this (these) metric(s) are appropriate on your data. That is, why is the metric appropriate for the task (e.g., in terms of the business case for the task). Please note: rarely is accuracy the best evaluation metric to use. Think deeply about an appropriate measure of performance.	15	11.5	This is a good start, but the false positives are different for each class. If you set this structure as a binary problem (either yes medal or not medal) then this argument works. But the multiclass problem you have currently needs more than just precision for each class. You need to interpret the confusion matrix directly.
Choose the method you will use for dividing your data into training and testing (i.e., are you using Stratified 10-fold cross validation? Why?). Explain why your chosen method is appropriate or use more than one method as appropriate. For example, if you are using time series data then you should be using continuous training and testing sets across time. Convince me that your cross validation method is a realistic mirroring of how an algorithm would be used in practice.	15	15	This is okay, but you probably do not need so many folds because you have a large amount of data.
Create a custom implementation of the multi-layer perceptron. Start with the implementation given to you in the course. Update the MLP class to: > When instantiated, use a selectable phi function for the first layer: either sigmoid or linear > Use a selectable cost function when instantiated: either quadratic or cross entropy >Add support for any number of hidden layers (user customizable).	20	18	I think you will need to initialize the weights in a better way (like Glorot normalization)--otherwise this will likely return "chance" for the output. > Yep. This is verified with the Iris dataset. It should have near 100% accuracy. > There might also be another bug somewhere--but I can't see it.
Tune the hyper-parameters of your MLP model (phi function, objective function, and number of layers). While tuning hyper-parameters, analyze the results using your chosen metric(s) of evaluation. **Visualize the evaluation metric(s) versus the hyper-parameters.** Conclude what combination of parameters are best.	15	15	Excellent for what you had.
Visualize the magnitude of the gradients in each layer of the neural network versus the training iteration. Do the gradients stay consistent in each layer?	10	0	Skipped.