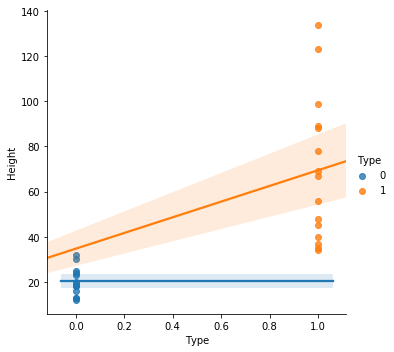
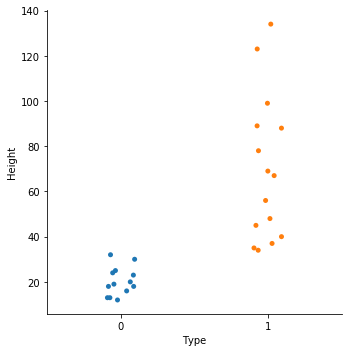
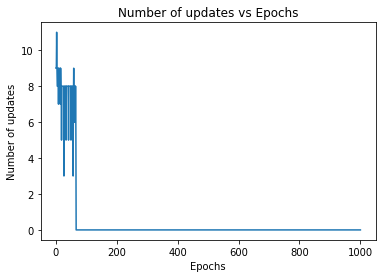
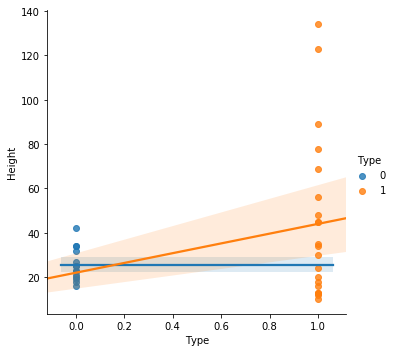
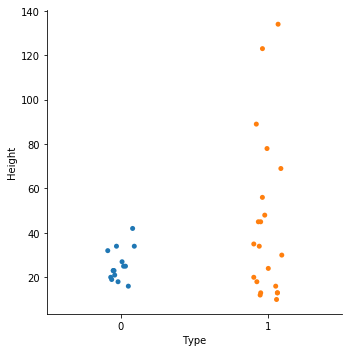
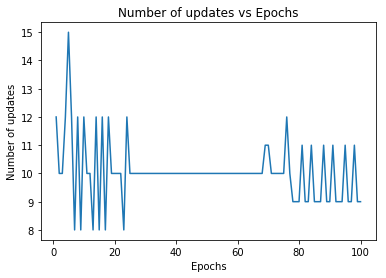
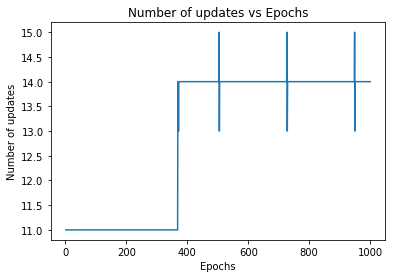
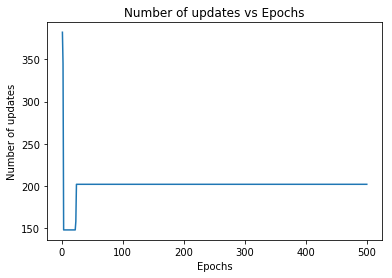
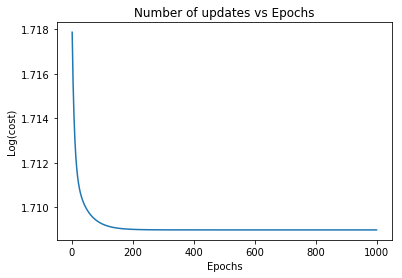
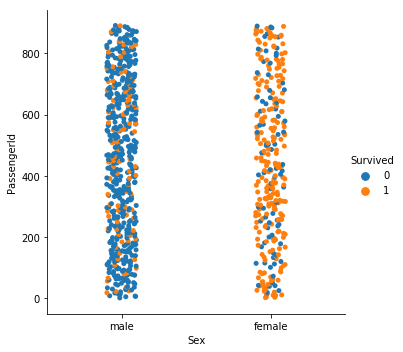
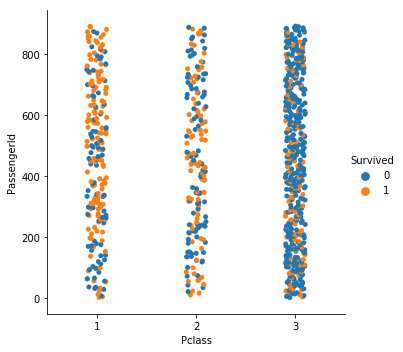
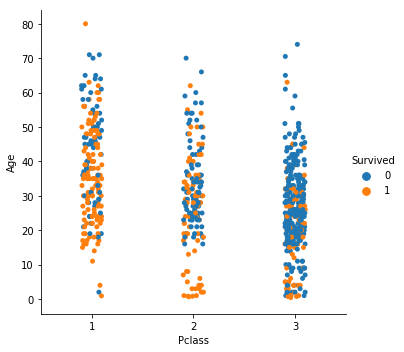
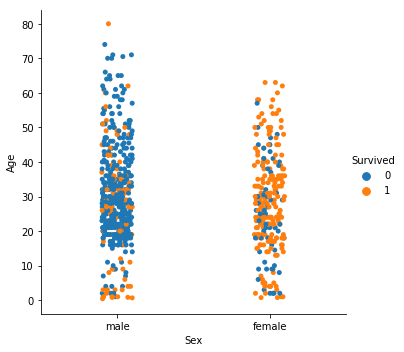
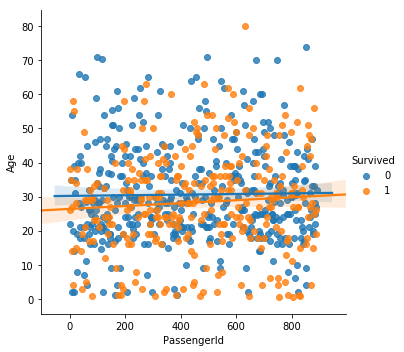
Components:

* stats() returns a cross tabulated table of passengers that survived based on Pclass and sex.
* scat() returns various scatter plots that display the correlation between various features vs the outcome of passengers.
* class Perceptron returns the accuracy of the perceptron.
* z\_input() returns the dot product of X vector and weight vector.
* weights() returns randomized weight vector composed of floats less than 1.
* predict() is step function which returns 1 if z is greater than or equal to 0, else returns 0.
* def fit() is activation function that measures the accuracy of the predictions and based off of the accuracy will then updated the weights to better predict the data. Also, returns the accuracy of the activation function, and if chosen, a list of the predictions.
* class TextBookAdalineGD is the Adaline algorithm provided by the textbook, used as a guide/check.
* testIt(testDat, testLabels) returns the accuracy of the test data label predictions.
* plotErrors() generates a updates vs epoch graph to display of convergence (if any) of the perceptron with the training data.
* class TitanicData() preps and analyses the titanic data set, returns numpy array of the test, and training data sets along with corresponding label arrays.
* def import\_clean\_TitanicData() cleans, preps and converts the titanic set data set into acceptable test, and training data sets to be feed into the perceptron algorithm.
* def LinSep\_data() creates a linearly separable data set, returns test and training data set; is later feed into the perceptron.
* def Not\_linSep\_data() creates a nonlinearly separable data set, returns test and training data set; is later feed into the perceptron.

Findings:

1. With a data set dogs which is linearly separable, there was a clear decision boundary line as a majority of the large dog species had a height above 30cm in comparison to the small dogs whom had heights between 10 < x < 30. The perceptron update rate converged with this linearly separable data.
   1. With learning rate of 0.001 and maximum epochs of 1000, my perceptron returned Accuracy 100.0 %
      1. 
      2. 
2. The nonlinear separable data composed of heights of cats and dogs is not linearly separable as cats have an average height from 20-30cm whereas dogs have an average height from 15-110cm. Thus, there is an overlap with species’ heights from 20-30cm. This is exemplified with the type vs height graphs as there is clearly an overlap of data and therefore no clear decision graph. Furthermore, the Updates vs. Epochs graphs demonstrates this data does not converge.
   1. With learning rate of 0.001 and maximum epoch of 100, my perceptron returns accuracy 27.27272727272727 %
      1. 
      2. 
   2. With learning rate 0.001 and epoch 1000
      1. 
3. Perceptron & AdalineSGD
   1. With my perceptron algorithm:
      1. With learning rate 0.003 and maximum epoch of 500, my perceptron returns 79.7752808988764 %
         1. 
   2. With Adaline stochastic gradient descent algorithm:
      1. With learning rate of 0.0001 and maximum epoch of 1000, my AdalineSGD returns accuracy 34.45692883895131 %
      2. 
   3. Note: np.random.shuffle does not maintain the probability distribution w/in the data throughout the data post array, therefore, the training and test can have different probabilities of survival; hence after each run my perceptron and AdalineSGD algorithms returned varying, but similar, accuracies.
4. Note: 1 indicates survived, 0 indicates death
   1. Graphs i and ii demonstrate a correlation between sex, and pclass with survival of passengers. As the diagrams exemplify, if passengers are female and in pclass 1 or 2, they were more likely to survive rather than their counterparts. Furthermore, those with the 1st and 2nd pclasses whom are adolescents/youth/young adults, were more likely to survive than the adults and elderly (practically within pclass 2). The most important feature within the data set is the sex as reducing the dimensionality of the data set from ['Pclass', 'Sex', 'Age', 'SibSp', 'Parch'] to [‘Sex’] increased the accuracy of the perceptron from 60-70% to about 80%. Therefore, in conclusion, sex is the most dominate and valuable predictive measure of passenger survival on the Titanic.
      1. 
      2. 
   2. 

0 1 female 3

male 77

2 female 6

male 91

3 female 72

male 300

1 1 female 91

male 45

2 female 70

male 17

3 female 72

male 47

Name: Survived, dtype: int64

Total passengers survived 342