**Assignment 2**

1. Using the model built in class for titanic dataset, calculate the accuracy of our model for passengers who survived and those didn't survive separately, i.e., accuracy for both possible outcomes.

For example:

+-----------+--------+----------+

**|passengerid|survived|prediction|**

+-----------+--------+----------+

| 1| 0.0| 0.0|

| 5| 0.0| 0.0|

| 8| 0.0| 0.0|

| 13| 0.0| 0.0|

Label Class 0 (Not survived): Accuracy 🡪 8/10 = 0.8

| 6| 0.0| 0.0|

| 14| 0.0| 0.0|

| 15| 0.0| 1.0|

| 7| 0.0| 0.0|

| 17| 0.0| 0.0|

| 19| 0.0| 1.0|

| 2| 1.0| 1.0|

| 11| 1.0| 1.0|

| 3| 1.0| 1.0|

| 12| 1.0| 1.0|

Label Class 1 (Survived): Accuracy 🡪 9/10 = 0.9

| 20| 1.0| 1.0|

| 4| 1.0| 1.0|

| 9| 1.0| 1.0|

| 16| 1.0| 1.0|

| 10| 1.0| 1.0|

| 18| 1.0| 0.0|

+-----------+--------+----------+

1. The 'fare' field in titanic dataset has continuous values in the range of zero to 512. This usually doesn't work well with most machine learning algorithms. Lets scale the fare values in the range of 0 to 1 and create a separate field (column) that has the scaled version of fare
2. We indexed the gender values in class exercise to zero or one (0 for male, 1 for female). A better practice is to convert such categorical fields to indicator fields. Create two additional fields in our titanic dataframe as:
   * IsMale: value would be 1 if the passenger was male otherwise 0
   * IsFemale: value would be 1 if the passenger was female otherwise 0
3. Build the model again with changes in 2 & 3 above, build the same logistic regression model and calculate the accuracy
4. We built and scored our model on the same data (t\_vec) which is not ideal. Split the dataset to use 70% of data for training and 30% for scoring. Calculate the accuracy on your test/scored data.