Name: __Alex Hansen_____

CSCI 303 Introduction to Data Science
Confusion Matrix Worksheet

1. What is a confusion matrix / what can it be used for? Predicts classification results and compares expected vs predicted outcomes. This allows you to determine how well the model can predict new data

	Predicted Label	
	Positive	Negative
Positive (True Label)	TP Hit	FN Miss
Negative (True Label)	FP False Alarm	TN Correct Reject

2. The diagonal elements are where the classifier predicted __correctly__; whereas, the off-diagonal elements are where the classifier _predicted incorrectly_.

3. Define the following terms/acronyms (and identify the one(s) we typically want to minimize):
   a. TP True Positive, correctly ID positive example (Hit)

   b. FP False Positive, ID'd a negative example as a positive (False Alarm)

   c. TN True Negative, correctly ID a negative example (Correct Reject)

   d. FN False Negative, ID'd a positive as a negative (Miss)

4. Fill in the diagram with:
   a. TP/FP/TN/FN
   b. Miss/Hit/False Alarm/Correct Reject

5. The performance measures we discussed are listed below. Provide the basic definition and formula for each of:
   a. Accuracy
   Percent of correctly classified observation. (TP+TN)/(Total Obsv)

   b. Precision
   Percent of positive predictions. (TP)/(TP+FP)

   c. Recall
   Percent of positive observations predicted as positive. (TP)/(TP+FN)

   d. F-score
   Balance between precision and recall. Harmon mean

   2(Prec x Recall)/(Prec + Recall)   or   (2TP)/(2TP + FP + FN)