CS-482 MACHINE LEARNING

HOMEWORK QUESTIONS

CHAPTER 1

1. Find a machine learning headline from either a) wired b) New York Times c) Independent d) IEEE e) ACM magazines f) Your favorite magazine, and, identify the following

a) What is the date and title of the headline?
b) What task was attempted (describe in 2 to 3 sentences)
c) What was the size of the data set(s) was used (how many, what kind etc- 2 to 3 sentences)
d) Is the technique(s) specified? If so, explain in 2 to 3 sentences
e) How successful was the model with the task? (Explain in 2 to 3 sentences)

2. The different types of machine learning problems. Determine whether the tasks de-scribed below involve supervised learning or unsupervised learning. For supervised learning problems, identify them as regression, classification, or probabilistic classification.

(a) Predict the risk of an accident at an intersection, given features such as the time of day and weather.
(b) Identify cars, bicyclists, and pedestrians in video taken by an autonomous vehicle's cameras.
(c) Determine the probability that there is a stop sign in an image.
(d) Generate new road scenarios (generate streets, place stop signs and intersections)for testing autonomous vehicles in a simulation

3. Train vs test datasets. Suppose you are building a classifier that identifies cats and dogs. You have a dataset of 3,000 images containing cats, dogs, or other objects (neither cat nor dog). You randomly split the data into a 2,500 image training set and a 500 image test set.

(a) Why is it important to "reserve" some images for the test dataset? (Why shouldn't we use all 3,000 images to train the classifier?)
(b) After training your classifier for a while, you observe it performs well on the training images, but poorly on the test images. What is one possible explanation?