**Assignment 1**

**Machine Learning**

MET CS 767

Faculty - Farshid Alizadeh-Shabdiz, PhD, MBA

**Assignment 1 (20 points)**

Under the phrase “Machine Learning” in Wikipedia, there is a section named “Applications”.

(The URL to the Application section is:

<https://en.wikipedia.org/wiki/Machine_learning#Applications> )

For 10 of the applications answer following questions:

* Write a sentence about the main application(s) of Machine learning in each and every application in the list.
* If you have to choose one, which of the learning methods (supervised, unsupervised or reinforcement learning) would you choose for each application. Explain in one sentence “why?”
* Between two big groups of ML methods, regression and classification, which one fits the best for each application? Explain in one sentence “why?”

**Assignment number 2 (20 points):**

Find two interesting websites about Machine Learning and/or Python on internet, for example following web sites:

ML: http://www.kaggle.com

ML: https://registary.opendata.aws

Python: <http://learnpython.org>/

Python: <https://docs.python.org/3/tutorial/>

* Write 3 most interesting things about each of the websites.
* Submit your answer on Blackboard, as you normally do for any assignment.

Also share your answer to this question with rest of the class by post your answer on Blackboard under “Discussion Forum” under “Week 1 Discussion”.

**Assignment number 3 (60 points):**

Predict median house value of California houses as a linear function of other parameters using California Housing dataset.

You can load California Housing dataset as follows:



You can learn about the data by looking at the DESCR method of the object.



1. **(10 points)** Plot histogram of numerical attributes of the data.
2. **(35 points)** Predict median price of the houses as a linear function of other parameters using Huber loss.

Use following Scikit-learn HuberRegressor. Keep all the default values as they are, except set the alpha to zero and set the epsilon to infinity.

*class*sklearn.linear\_model.**HuberRegressor**

Split data to training and test set (70, 30) using following function.

from sklearn.cross\_validation import train\_test\_split

Comment on the results.

1. **(15 points)** Repeat item 2 with epsilon set to one. Comment on the results.

