**Homework 2**

**Implementing logistic regression**

1. Download the election\_dataset.csv (This is the 1996 US President election dataset. We are going to predict vote of people based on some of their features!)
2. Open the Binary logistic regression-HW.ipynb to start developing your code. (Good news: I provided most functions you need!)
3. Split your dataset to test sets and training sets (80% training set, 20% test set).
4. Take Party Identification of the person, age, and the number of time person views the TV-news as the features for your prediction.
5. Train your model with the training dataset.
6. Print the accuracy of your model on the test data set.
7. Predict and print the predicted vote of a 25 years old weak democrat person who watch TV-news 7 times a week.
8. At the end, you should send me one Jupyter notebook (.ipynb file) which has just **one cell**, and by running that single cell, it should print accuracy of your model and predicted vote of the mentioned guy!

\*Hint: I have uploaded the implementation of the decision tree algorithm in Python in course material section. As steps are very similar, you can take advantage of that!

* **How to turn in your HW:** Email me your “Your Full Name.ipynb” file at [mjafa003@ucr.edu](mailto:mjafa003@ucr.edu).
* **Subject of your email should be** : “Homework 2 ML-Your Full Name “
* **Deadline:** Tuesday July 17 @ 3:00 pm.

\*Here is the description of each column of the dataset:

1. popul: means the population in the census place in 1000
2. TVnews: the number of time the voter views the Tv news in a week.
3. selfLR: is the person’s self–reported political learnings from left to right.
4. ClinLR: is the person’s impression on Bill Clinton’s Political learning from left to right
5. DoleLR: is the person impression of Bob Dole’s Political learnings from left to right.
6. PID: party Identification of the person. If the PID is
   * + 0 means the Strong Democrat,
     + 1 means Week democrat,
     + 2 means Independent democrat likewise
7. Age: age of the voter.
8. Educ: education qualification of the voter.
9. Income: income of the voter.
10. Vote: the vote is the target which we are going to predict using the trained logistic regression model. Vote having two possible outcomes: 0 means Clinton, 1 means Dole