Homework 3:  
Create a logistic regression model using a dataset of your choice.  
Run cross validation for different values of your regularization parameter.  
  
Extra: Perform K-means clustering on a dataset of your choice. Compare results using different similarity metrics.

Ipython:

<http://nbviewer.ipython.org/urls/raw.github.com/tomoeyukishiro/GA_homework/master/Day_5_Logistic_Regression.ipynb>

Data used in tutorial:

<https://github.com/tomoeyukishiro/GA_homework/blob/master/data/test.csv>

SELECT a.activity\_id, count(a.activity\_id) AS num\_actions, b.goal,

SUM(CASE WHEN a.recruiter\_id IS NOT NULL THEN 1 ELSE 0 END) AS num\_recruited\_actions,

-- SUM(CASE WHEN c.context\_id = a.activity\_id THEN 1 ELSE 0 END) AS num\_posts,

b.created\_at

FROM action\_credit\_silo.action\_credits AS a

JOIN descolada\_prod.activities AS b ON (a.activity\_id = b.id)

-- LEFT JOIN descolada\_prod.post\_references AS c ON (b.id = c.context\_id)

JOIN descolada\_prod.campaigns AS d ON (b.campaign\_id = d.id)

WHERE d.id > 82

AND d.launched\_at IS NOT NULL

AND d.deleted\_at IS NULL

AND d.title NOT LIKE '%test%'

AND d.state NOT LIKE 'Inactive'

-- AND c.context\_type = 'Activity'

GROUP BY activity\_id

Copy data from my local desktop to the aws node

scp Desktop/hw3\_train.csv [GA2@54.200.59.194:~/GA\_Homework/hw3](mailto:GA2@54.200.59.194:~/GA_Homework/hw3)

scp Desktop/hw3\_test.csv [GA2@54.200.59.194:~/GA\_Homework/hw3](mailto:GA2@54.200.59.194:~/GA_Homework/hw3)

Copy .py file to the node, too

scp GA\_github/GA\_Homework/hw3/hw3.py [GA2@54.200.59.194:~/GA\_Homework/hw3](mailto:GA2@54.200.59.194:~/GA_Homework/hw3)

Now log into aws node

ssh [GA2@54.200.59.194](mailto:GA2@54.200.59.194)

python

Run hw2.py script (line by line in Python to check that it’s working as I go, or I can write an entire script and run it using ‘python hw2.py’

To push to git

cd GA\_github/GA\_Homework/hw2/

ls

git status

git add hw2.py

git commit

git push

And then check github.com to make sure it’s there!

Could take some percentage of goal as good or bad.