Lecture 5:

Quiz #2 (Fri – Sunday)

Assignment 5

Part 1:

* linear separation for labels

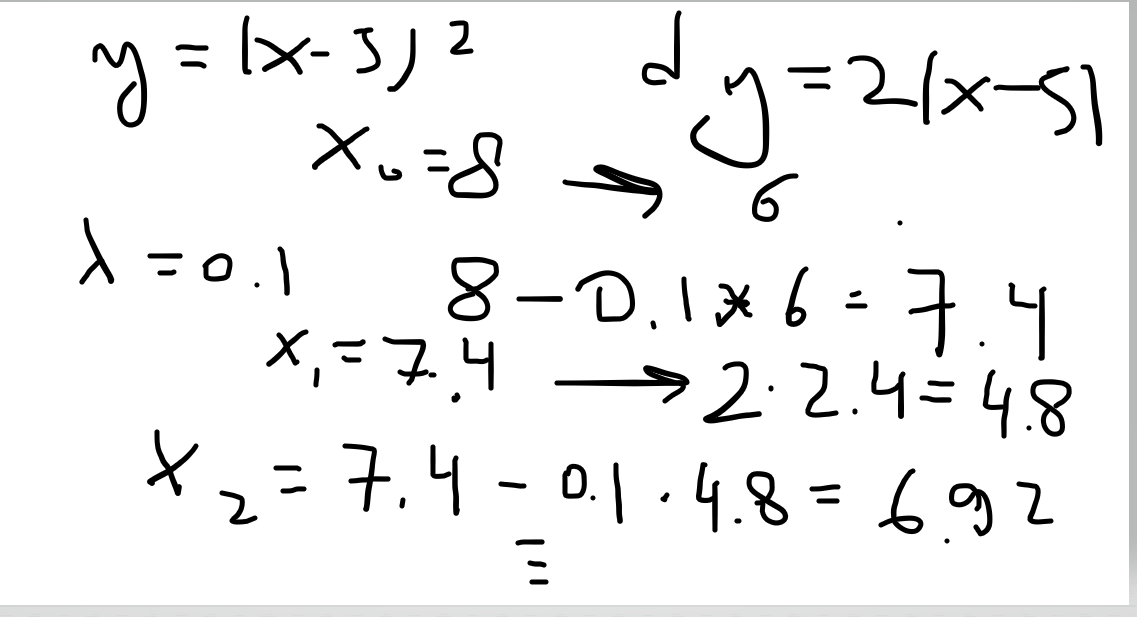
(manually remove my points

* kNN classifier for your stock
* logistic regression

Part 2:

* kNN and logistic classifiers to detect fake banknotes

On Gradient descent:



On scaling:

A = (100, 1)

B = (110, 1)

C = (100, 2)

Q: is A closer to B or A is closer to C

d(A,B)\*\*2 = (110 – 100)\*82 + (1-1)\*\*2 = 20\*= 10\*\*2

distance is 10

d(A, c) = 1

How do we scale:

3,4,5,6

Min-Max scaling

max = 6

min = 3

x ------------------🡪 (x –min)/(max-min)

3,4,5,6 ---------🡪 0, 1/3, 2/3, 1

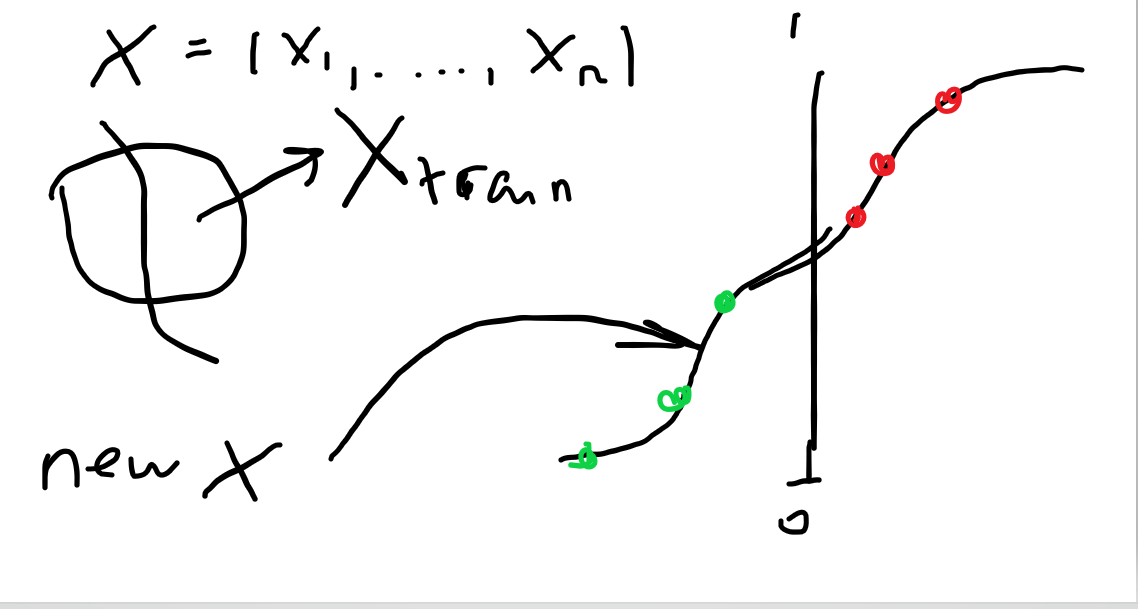
standard scaling:

{x} ---------------------🡪 mu, sigma

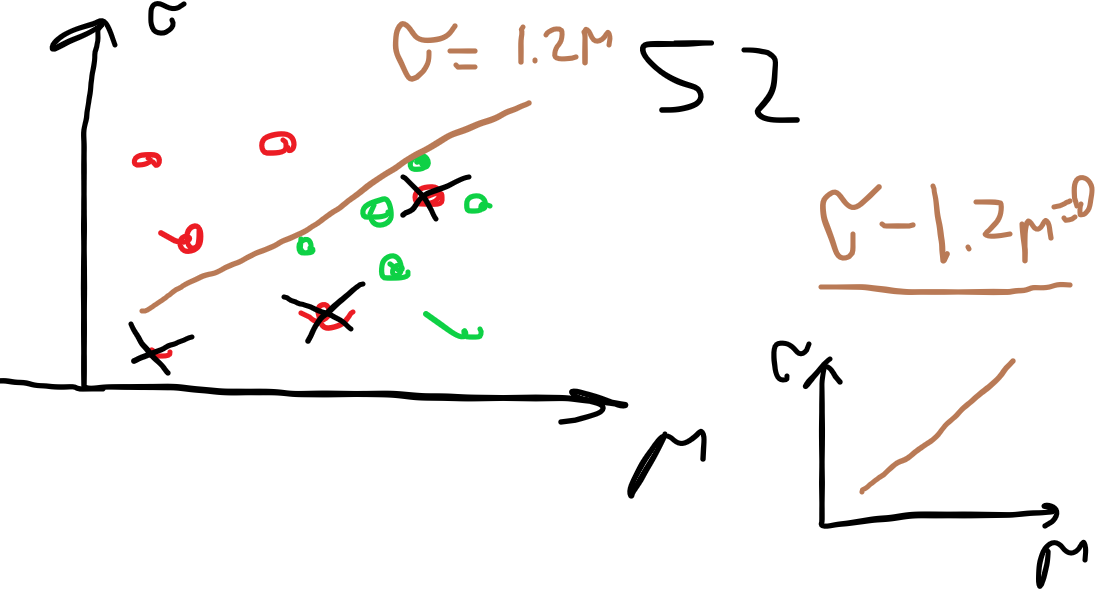
x ------------------🡪 (x-mu)/sigma

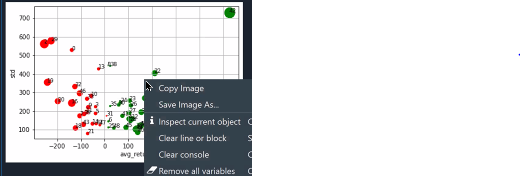
Logistic regression:

given points with labels, compute “weights” so that points in X\_train are located on the   
“logit” function. Then for any new point x\*, find out where is it located on the login function. Classify this point depending whether its logit value is greater or smaller than 1/2



Homework notes





For question 4

kNN ---------------🡪 accuracy A\*

with optimal k\* (if you use all 4 features)

take f2, f3, f4 (drop f1) ---🡪 A1

f1, f3, f4 (drop f2) ---🡪 A2 (accuracy)

f1, f2, f4 (drop f3) ---🡪 A3

f1, f2,f3 (drop f4) ---🡪 A4

(use the same k\* that you computed before)

each time time you change your features, split the dataset into X\_train and X\_test

Question on BUID

U12345678

f1 = 5, f2 = 6, f3=7, f4 = 8