Data Science: Deep Learning Prerequisites – Linear Reg in Python Notes

**Section 3: Multiple Linear Regression and Polynomical Regression Notes**

multiple linear regression

if there are mulitiple inputs int eh data set

most realistic – usually multiple inputs affect the outpout

housing price and other things

for simple linear regression

it is (x1,y1) etc….

for multiple linear regression

still have this format but x1 is now a fector, not a scalar

dimensionality is equal to the size of X represented by letter D

This is just like the notes from Stats class

dimmentionality is the (number of input variables -1)

model is:

y^=wtx+b

wtx means “w transpose x”

transpose is

<https://www.khanacademy.org/math/linear-algebra/matrix-transformations/matrix-transpose/v/linear-algebra-transpose-of-a-vector>

transpose becomes X times the w vector

so See notes to explain how this is like a dot product

can absorb b into w by appending 1 to vector x in cell 0 and setting w0 to b

y^=W0X0+W1X1+W2X­2…… where W0=b and X0=1

things to note about Data matrix X

it is an NxD matrix

N=number of samples

D=numper of inputs and features

1 row of X represents one full sample

this is called the feature vector

need to transpose W so that we can multiply since you need to have a valid matrix vector

you can actually just multiply the matricies

because 4x3 matrix can be multiplied by a 3x1 matrix

so don’t separate out ingto separate matrices

just multiply all

after finding predictions we need to minimize error

so we use our error function but plug in our new prediction matrix

**see page 4 of notes**

**See the example in the notes on page 3**

**Linear regression in python**

this isa lecture

ohm’s law V=IR is a linear equation

V=vol tate

I=current

R=resistance

Current (I) is the independent variable

given a set of inputs and outputs you can plot on graph

creates an almost perfect line

always some element of error

drawing the line of best fit

try to calculate the slope of the lin

R=slope=V/I

linear can apply to anything

blood pressure versus age

blood pressure vs weight

All data is the same

there is a systematic way to find line of best fitSimple linear regression

this is yi=axi+b

y*i*is expected y

xi is the input ex

we want the line of best fit to be as close to possible (minimizing error)

finding differences does not work

errors of +5 and -5 = 0

Step 1: need to find the “sum of squared errors”

Error (E)=summation (yi-yhati)2

Step 2: take the derivive of Error function to minimize the result

example

E=0.5t2-t

E`=t-1

so by setting derivative equal to 0

0=t-1

t=1

in our problem

Error (E)=summation (yi-yhati)2

So pluggin in our line of best fit function

E=summation(yi-(axi+b))2 or E=summation(yi-axi-b)2

yi and xi are given due to datapoints

have to take partial derivatives since multiple variables <https://www.youtube.com/watch?v=SbfRDBmyAMI>

<https://www.google.com/search?q=finding+derivatives+polynomials&ie=utf-8&oe=utf-8&client=firefox-b-1-ab#kpvalbx=1>

<https://www.youtube.com/watch?v=-_8DFxTl0Ls>

<http://polisci.msu.edu/jacoby/msu/ppl801/handouts/Handout,%20Summations,%202010.pdf>

Need to use the chain rule on the original ufnction to get the derivative

<https://www.khanacademy.org/math/ap-calculus-ab/ab-derivative-rules/ab-chain-rule/v/chain-rule-introduction>

Steps:

1. take derivative of outside quantity

2. need to then mulitpily result by derivative of interior

example

h(x)=(sinx)2

h`(x)=(dh/dx)=2(sinx)(derivative of sign x)

look at notes to show that you can convert

sum(xi)=X bar (if divided by N)

dot product resources:

<https://en.wikipedia.org/wiki/Dot_product>

**see notes sheet**

**Define the problem derive the solution**

process

1. define the problem

start with training samples that resemble a line

inputs (x1, x2 x3 x4)

outputs (y1, y2, y3,…)

line defined in yhat=ax+b

**Summary:**

In theory (this may just be the proof)

1. find the equation of the sum of square derrors

2. take partial derivatives of a and b and set to zero, then solve to get minimum a and b values

3. those create the line of best fit

4. next steps: get line of best fit

**Exercise: Theory Versus Coding**

the code helps conceptualize the results

**R-Squared Quiz:**

1. What leads to a negative R2?

we predict worse than mean of the target values

it would have been better to predict just the mean of the target values