

Linear Regression

October 12, 2015

```
In [1]: import numpy as np
In [2]: import urllib
In [3]: import scipy.optimize
In [25]: import random
         from sklearn import datasets, linear_model, metrics
In [5]: def parseData(fname):
         for l in urllib.urlopen(fname):
             yield eval(l)
In [6]: print "Reading data..."
         data = list(parseData("file:///Users/YW/Dropbox/UCSD/FA2015/CSE255/HW/hw1/beer_50000.json"))
         print "done"

Reading data...
done
In [7]: def feature(datum):
         feat = [1]
         feat.append(datum['beer/ABV'])
         return feat
In [8]: X = [feature(d) for d in data]
In [9]: y = [d['review/taste'] for d in data]
In [10]: theta, residuals, rank, s = np.linalg.lstsq(X, y)
In [11]: print 'Theta0:%s, Theta1:%s'%(theta[0], theta[1])

Theta0:3.11521115232, Theta1:0.109055074172
In [37]: X1 = np.matrix(X)
         y1 = np.matrix(y)
         print np.linalg.inv(X1.T * X1) * X1.T * y1.T

[[ 3.11521115]
 [ 0.10905507]]
In [79]: def feature5(datum):
         feat = [1]
         feat.append(datum['beer/ABV'])
         feat.append(datum['beer/ABV']**2)
         feat.append(datum['beer/ABV']**3)
         feat.append(datum['beer/ABV']**4)
         feat.append(datum['beer/ABV']**5)
         return feat
```

```

In [80]: X5 = [feature5(d) for d in data]

In [81]: y5 = [d['review/taste'] for d in data]

In [82]: theta, residuals, rank, s = np.linalg.lstsq(X5, y5)

In [83]: print theta, residuals, rank, s

[ 1.40998006e+00  6.71979699e-01 -5.57829661e-02  1.95198865e-03
 -3.03848616e-05  1.73376523e-07] [ 21821.27545336] 6 [ 6.93445043e+08  2.17982482e+06  7.70108652e+06
 9.97195416e+01  5.73256063e+00]

In [84]: regr = linear_model.LinearRegression(fit_intercept=False)

In [85]: regr.fit(X5, y5)

Out[85]: LinearRegression(copy_X=True, fit_intercept=False, n_jobs=1, normalize=False)

In [86]: print 'Coefficients:%s\n'% regr.coef_

Coefficients: [ 1.40998006e+00  6.71979699e-01 -5.57829661e-02  1.95198865e-03
 -3.03848616e-05  1.73376523e-07]

In [87]: print 'MSE: %s' % np.mean((regr.predict(X5) - y5) ** 2)
          print 'MSE: %.5f' % metrics.mean_squared_error(y5, regr.predict(X5), sample_weight=None)

MSE: 0.436425509067
MSE: 0.43643

In [88]: train = data[:25000]
          test = data[25000:]

In [89]: def features(datum, degree):
          feat = [1]
          for i in xrange(1, degree+1):
              feat.append(datum['beer/ABV']**i)
          return feat

In [101]: lastTestingError = 0

In [114]: deg = 1

In [115]: while True:
            regr = linear_model.LinearRegression(fit_intercept=False)
            X = [features(d, deg) for d in train]
            y = [d['review/taste'] for d in train]
            X_test = [features(d, deg) for d in test]
            y_test = [d['review/taste'] for d in test]
            regr.fit(X, y)
            thisTestingError = np.mean((regr.predict(X_test) - y_test) ** 2)
            thisTrainError = np.mean((regr.predict(X)-y) ** 2)
            print "Train:", thisTrainError
            print "Test:", thisTestingError
            print "coef:", regr.coef_
            print len(X[0])
            if np.fabs(thisTestingError-lastTestingError) < 0.000001:
                break
            deg = deg + 1
            lastTestingError = thisTestingError

```

Train: 0.483983105115
 Test: 0.423776528023
 coef: [2.99503282 0.11690802]
 2
 Train: 0.471743067557
 Test: 0.427256126
 coef: [2.62007309 0.20716481 -0.00496806]
 3
 Train: 0.457832195847
 Test: 0.432820707084
 coef: [1.57847740e+00 5.19869113e-01 -2.97470415e-02 3.97626061e-04]
 4
 Train: 0.451641221709
 Test: 0.438301732573
 coef: [7.17629022e-01 8.26164532e-01 -6.32189636e-02 1.67076516e-03
 -1.40075070e-05]
 5
 Train: 0.451335575196
 Test: 0.439820774965
 coef: [1.16389773e+00 6.12905604e-01 -2.85943943e-02 -6.28209940e-04
 4.41325901e-05 -4.79191916e-07]
 6
 Train: 0.450769302041
 Test: 0.443012263638
 coef: [1.94901771e+00 1.79935890e-01 5.75772252e-02 -8.45991886e-03
 3.82283840e-04 -7.06992367e-06 4.63100297e-08]
 7
 Train: 0.45074188066
 Test: 0.442575342636
 coef: [1.70847546e+00 3.47096902e-01 1.39022021e-02 -2.95198356e-03
 2.49609809e-05 4.72513678e-06 -1.40946836e-07 1.12958340e-09]
 8
 Train: 0.450231148043
 Test: 0.442487701726
 coef: [3.93690397e-01 1.40863997e+00 -3.13749111e-01 4.74923395e-02
 -4.20781201e-03 2.03296270e-04 -5.28225580e-06 6.93371692e-08
 -3.60135053e-10]
 9
 Train: 0.449716240495
 Test: 0.439253459889
 coef: [6.57779065e-01 7.66518411e-01 4.06647965e-02 -3.84794380e-02
 6.77043490e-03 -5.85586427e-04 2.71796542e-05 -6.83946306e-07
 8.76795980e-09 -4.47490401e-11]
 10
 Train: 0.485239771582
 Test: 0.457827607951
 coef: [8.10035383e-03 2.95349476e-02 7.80426486e-02 1.25171126e-01
 -4.56697918e-02 6.52647339e-03 -4.84697958e-04 2.00717535e-05
 -4.64486475e-07 5.59715969e-09 -2.72729507e-11]
 11
 Train: 0.627028248862
 Test: 0.646353376223
 coef: [3.13393784e-04 2.12106023e-03 5.06678428e-03 1.51375880e-02
 2.59611730e-02 -9.65927443e-03 1.37038921e-03 -1.00082775e-04

4.07376218e-06 -9.28854537e-08 1.10590758e-09 -5.33743373e-12]

12
Train: 2.85408190325
Test: 4.53394584743
coef: [1.20060415e-08 -1.67445314e-06 5.25536588e-07 3.21145550e-06
1.74858508e-05 7.57760222e-05 1.95196988e-04 -5.09012896e-05
4.96757208e-06 -2.38620849e-07 6.03395828e-09 -7.69501605e-11
3.89503838e-13]

13
Train: 8.18203669655
Test: 10.0715024042
coef: [3.95168107e-14 1.06781904e-10 3.10753427e-12 3.01578879e-11
2.67364021e-10 2.24185996e-09 1.68928241e-08 1.03194964e-07
3.87685919e-07 -6.75444710e-08 4.24511502e-09 -1.25235102e-10
1.76192140e-12 -9.53340770e-15]

14
Train: 9.554158671
Test: 11.0448045432
coef: [2.16639158e-16 -5.08056120e-11 -2.29606780e-14 1.97454960e-13
1.89012696e-12 1.75238256e-11 1.52805342e-10 1.19077076e-09
7.48268931e-09 2.87797828e-08 -4.94168737e-09 3.08053486e-10
-9.04048899e-12 1.26734339e-13 -6.83975029e-16]

15
Train: 10.6493250888
Test: 11.7727856209
coef: [1.17714518e-18 -4.45631245e-13 -4.58771397e-15 1.23856013e-15
1.24714430e-14 1.23881558e-13 1.18594897e-12 1.06296521e-11
8.47778730e-11 5.43078453e-10 2.12212634e-09 -3.59970437e-10
2.22832356e-11 -6.50993035e-13 9.09743964e-15 -4.89877188e-17]

16
Train: 14.658067654
Test: 14.6303906263
coef: [2.58346426e-25 -1.12064050e-15 -2.07631352e-19 6.14616717e-22
5.33209390e-21 6.67769289e-20 8.47113777e-19 1.08018529e-17
1.36403547e-16 1.65985013e-15 1.85096217e-14 1.70915296e-13
9.99592835e-13 -9.93492947e-14 3.61772635e-15 -5.74964411e-17
3.36561166e-19]

17
Train: 15.0695872189
Test: 15.0707338786
coef: [8.24007104e-28 1.23742293e-16 -3.04283198e-21 -3.21208823e-23
1.84267302e-23 2.36419159e-22 3.08338023e-21 4.06979631e-20
5.38745687e-19 7.03731870e-18 8.81727942e-17 1.00719518e-15
9.47871416e-15 5.62466260e-14 -5.58960316e-15 2.03534267e-16
-3.23472985e-18 1.89347485e-20]

18
Train: 15.3488085966
Test: 15.3919509124
coef: [2.42412870e-30 -7.77039051e-18 4.74890909e-21 1.43152594e-25
6.09204213e-26 7.72181044e-25 1.03250682e-23 1.40202437e-22
1.92188975e-21 2.63248076e-20 3.54182580e-19 4.54793549e-18
5.29779276e-17 5.06144231e-16 3.03755177e-15 -3.01830340e-16
1.09902445e-17 -1.74663973e-19 1.02240519e-21]

19

```

Train: 15.8358242625
Test: 15.9550205998
coef: [ 1.09721348e-44 -3.80173473e-28 -1.68836187e-32 -5.46659687e-36
1.45912708e-38 2.64070845e-37 7.73757971e-36 2.23663703e-34
6.35375978e-33 1.76457920e-31 4.75650672e-30 1.23161080e-28
3.01607501e-27 6.81360469e-26 1.35910058e-24 2.18837256e-23
2.21986060e-22 -1.46300726e-23 3.16900977e-25 -2.25541643e-27]

```

20

```

Train: 15.8358344939
Test: 15.9550264602
coef: [ 1.16144638e-47 8.03196982e-30 8.59344924e-34 1.28367662e-36
-2.96853875e-39 2.89867746e-40 8.58237099e-39 2.51480251e-37
7.26966628e-36 2.06518724e-34 5.73554315e-33 1.54605131e-31
4.00322961e-30 9.80346876e-29 2.21469970e-27 4.41763235e-26
7.11310207e-25 7.21543628e-24 -4.75984657e-25 1.03181543e-26
-7.34803958e-29]

```

21

```

Train: 15.8358386458
Test: 15.9550286554
coef: [ 1.23480922e-50 1.43384821e-30 -5.93199965e-36 1.78108556e-38
-5.26064528e-41 3.13096731e-43 9.43944138e-42 2.79559030e-40
8.19179384e-39 2.36806648e-37 6.72730260e-36 1.86834530e-34
5.03624712e-33 1.30404895e-31 3.19347321e-30 7.21436936e-29
1.43904051e-27 2.31708659e-26 2.35041756e-25 -1.55172618e-26
3.36590504e-28 -2.39825616e-30]

```

22

```

Train: 15.8358403672
Test: 15.955029486
coef: [ 1.31263710e-53 -5.38892493e-32 -3.23662641e-36 3.15709069e-39
2.62616297e-42 4.17109278e-46 1.02960295e-44 3.07504109e-43
9.10714282e-42 2.66863746e-40 7.71445794e-39 2.19155799e-37
6.08652461e-36 1.64066299e-34 4.24821298e-33 1.04034086e-31
2.35023179e-30 4.68797439e-29 7.54838817e-28 7.65695804e-27
-5.05828127e-28 1.09778541e-29 -7.82520950e-32]

```

23

```

In [108]: print X[0]
          print X[1]
          print regr.coef_

```

```

[1, 5.0, 25.0, 125.0, 625.0, 3125.0, 15625.0, 78125.0, 390625.0, 1953125.0, 9765625.0, 48828125.0, 244140625.0, 1220703125.0, 6103515625.0, 30517578125.0, 152587890625.0, 762939453125.0, 3814697265625.0, 19073486328125.0, 95367431640625.0, 476837158203125.0, 2384185791015625.0, 11920928955078125.0, 59604644775390625.0, 298023223876953125.0, 1490116119384765625.0, 7450580596923828125.0, 37252902984619140625.0, 186264514923095703125.0, 931322574615478515625.0, 4656612873077392578125.0, 23283064365386962890625.0, 116415321826934814453125.0, 582076609134674072265625.0, 2910383045673370361328125.0, 14551915228366851806640625.0, 72759576141834259033203125.0, 363797880709171295166015625.0, 1818989403545856475830078125.0, 9094947017729282379150390625.0, 45474735088646411895751953125.0, 227373675443232059478759765625.0, 1136868377216160297393798828125.0, 5684341886080801486968994140625.0, 28421709430404007434844970703125.0, 142108547152020037174224853515625.0, 710542735760100185871124267578125.0, 3552713678800500929355621337890625.0, 17763568394002504646778106689453125.0, 88817841970012523233890533447265625.0, 444089209850062616169452667236328125.0, 2220446049250313080847263336181640625.0, 11102230246251565404236316680908203125.0, 55511151231257827021181583404541015625.0, 277555756156289135105907917022705078125.0, 1387778780781445675529539585113525390625.0, 6938893903907228377647697925567626953125.0, 34694469519536141888238489627838134765625.0, 173472347597680709441192448139190673828125.0, 867361737988403547205962240695953369140625.0, 4336808689942017736029811203479766845703125.0, 21684043449710088680149056017398834228515625.0, 108420217248550443400745280086994171142578125.0, 542101086242752217003726400434970855712890625.0, 2710505431213761085018632002174854278564453125.0, 13552527156068805425093160010874271392822265625.0, 67762635780344027125465800054371356464111328125.0, 338813178901720135627329000271856782320556640625.0, 1694065894508600678136645001359283911602783203125.0, 8470329472543003390683225006796419558013916015625.0, 42351647362715016953416125033982097790069580078125.0, 211758236813575084767080625169910488950347900390625.0, 1058791184067875423835403125849552444751739501953125.0, 5293955920339377119177015629247762223758697509765625.0, 26469779601696885595885078146238811118793487548828125.0, 132348898008484427979425390731194055593967437744140625.0, 661744490042422139897126953655970277969837188720703125.0, 3308722450212110699485634768279851389849185943603515625.0, 16543612251060553497428173841399256949245929718017578125.0, 82718061255302767487140869206996284746229648590087890625.0, 413590306276513837435704346034981423731148242950439453125.0, 2067951531382569187178521730174907118655741214752197265625.0, 10339757656912845935892608650874535593278706073760986328125.0, 51698788284564229679463043254372677966393530368804931640625.0, 258493941422821148397315216271863389831967651844024658203125.0, 1292469707114105741986576081359316949159838259220123291015625.0, 6462348535570528709932880406796584745799191296100616455078125.0, 32311742677852643549664402033982923728995956480503082275390625.0, 161558713389263217748322010169914618644979782402515411376953125.0, 807793566946316088741610050849573093224898912012577056884765625.0, 4038967834731580443708050254247865466124494560062885284423828125.0, 20194839173657902218540251271239327330622472800314426422119140625.0, 100974195868289511092701256356196636653112364001572132110595703125.0, 504870979341447555463506281780983183265561820007860660552978515625.0, 2524354896707237777317531408904915916327809100039303302764892578125.0, 12621774483536188886587657044524579581639045500196516513824462890625.0, 63108872417680944432938285222622897908195227500982582569122314453125.0, 315544362088404722164691426113114489540976137504912912845611572265625.0, 1577721810442023610823457130565572447704880687524564564228057861328125.0, 7888609052210118054117285652827862238524403437622822821140289306640625.0, 39443045261050590270586428264139311192622017188114114105701446533203125.0, 197215226305252951352932141320696555963110085940570570528507232666015625.0, 986076131526264756764660706603482779815550429702852852642536163330078125.0, 4930380657631323783823303533017413899077752148514264263212680816650390625.0, 24651903288156618919116517665087069495388760742571321316063404083251953125.0, 123259516440783094595582588325435347476943803712856606580317020416259765625.0, 616297582203915472977912941627176737384719018564283032901585102081298778125.0, 3081487911019577364889564708135883686923595092821415164507925510406493890625.0, 15407439555097886824447823540679418434617975464107075822539627552032469453125.0, 77037197775489434122239117703397092173089877320535379112698137760162347265625.0, 385185988877447170611195588516985460865449386602676895563490688800811736328125.0, 1925929944387235853055977942584927304327246933013384477817453444004058681640625.0, 9629649721936179265279889712924636521636234665066922389087267220020293408203125.0, 48148248609680896326399448564623182608181173325334611945436336100101467041015625.0, 240741243048404481631997242823115913040905866626673059727181680500507335205078125.0, 1203706215242022408159986214115579565204529333133365298635908402502536676025390625.0, 6018531076210112040799931070577897826022646665666826493179542012512683380126953125.0, 30092655381050560203999655352889489130113233328334132465897710062563416900634765625.0, 150463276905252801019998276764447445650566166641670662329488550312817084503173828125.0, 752316384526264005099991383822237228252830833208353311647442751564085422515869140625.0, 3761581922631320025499956919111186141264154166041766558237213757820427112579345703125.0, 18807909613156600127499784595555930706320770830208832791186068789102135562896728515625.0, 94039548065783000637498922977779653531603854151044163955930343945510677814483642578125.0, 470197740328915003187494614888898267658019270755220819779651719727553389072418212890625.0, 2350988701644575015937473074444491338290096353776104098898258598637766945362091064453125.0, 11754943508222875079687365372222456691450481768880520494491292993188834726810455322265625.0, 58774717541114375398436826861112283457252408844402602472456464965944173634052276611328125.0, 293873587705571876992184134305561417286262044222013012362282324829720868170261383056640625.0, 1469367938527859384960920671527807086431310221110065061811411624148604340851306915283203125.0, 7346839692639296924804603357639035432156551105550325309057058120743021704256534576416015625.0, 36734198463196484624023016788195177160782755527751626545285290603715108521282672882080078125.0, 183670992315982423120115083940975885803913777638758132726426453018575542606413364410400390625.0, 918354961579912115600575419704879429019568888193790663632132265092877713032066822052001953125.0, 4591774807899560578002877098524397145097844440968953318160661325464388565160334110260009765625.0, 22958874039497802890014385492621985725489222204844766590803306627321942825801670551300048828125.0, 114794370197489014450071927463109928627446111024223832954016533136609714129008352756500244140625.0, 5739718509874450722503596373155496431372305551211191647700826656830485706450417637825012207203125.0, 28698592549372253612517981865777482156861527756055958238504133284152428532252088189125061036015625.0, 143492962746861268062589909328887410784307638780279791192520666420762142661260440945625305180078125.0, 717464813734306340312949546644437053921538193901398955962603332103810713306302204728126525900390625.0, 3587324068671531701564747733222185269607690969506994779813016660519053566531511023640632629501953125.0, 17936620343357658507823738666110926348038454847534973899065083302595267832657555118203163147509765625.0, 89683101716788292539118693330554631740192274237674869495325416512976339163287775591015815737548828125.0, 448415508583941462695593466652773158700961371188374347476627082564881695816438877955079078687744140625.0, 2242077542919707313477967333263865793504806855941871737383135412824408479082194389775395393438720703125.0, 11210387714598536567389836666319328967524034279709358686915677064122042395410971948876976967193603515625.0, 56051938572992682836949183331596644837620171398546793434578385320610211977054859744384884835968017578125.0, 280259692864963414184745916657983224188100856992733967172891926603051059885274298721924424179840087890625.0, 1401298464324817070923729583289916120940504284963669835864459633015255299426371493609622120899200439453125.0, 7006492321624085354618647916449580604702521424818349179322298165076276497131857468048110604496002197265625.0, 35032461608120426773093239582247903023512607124091745896611490825381382485659287340240553022480010986328125.0, 175162308040602133865466197911239515117563035620458729483057454126906912428296436701202765112400054931640625.0, 875811540203010669327330989556197575587815178102293647415287270634534562114282183506013825562000274658203125.0, 4379057701015053346636654947780987877939075890511468237076436353172672810571410917530069127810001373291015625.0, 21895288505075266733183274738904939389695379452557341185382181765863364052857054587650345639050006866455078125.0, 109476442525376333665916373694524696948476897262786705926910908829316820264285272937751728195250034332275390625.0, 547382212626881668329581868472623484742384488311393529634554544146584101321426364688758640976250171661376953125.0, 2736911063134408341647909342363117423711922441556967648172772720732920506607131823443793204881250858306884765625.0, 136845553156720417082395467118155871185596122077848382408638636036646025330356591172189660244062504291534423828125.0, 684227765783602085411977335590779355927980610389241912043193180183230126651782955860948301220312502145672119140625.0, 3421138828918010427059886677953896779639903051946209560215965900916150633258914779304741506101562501072836095703125.0, 17105694144590052135299433389769483898199515259731047801079829504580753166294573896523707530507812500536430478515625.0, 85528470722950260676497166948847419490997576298655239005399147522903765831472869482618537652539062500268202392578125.0, 427642353614751303382485834744237097454987881493276195026995737614518829157364347413092688262695312500134111962890625.0, 21382117680737565169124291737211854872749394074663809751349786880725941457868217370654634413134765625000670598114453125.0, 106910588403687825845621458686059274363746970373319048756748934403629707289341086853273172065723828125000335279572265625.0, 534552942018439129228107293430296371818734851866595243
```

In []: