Homework 2
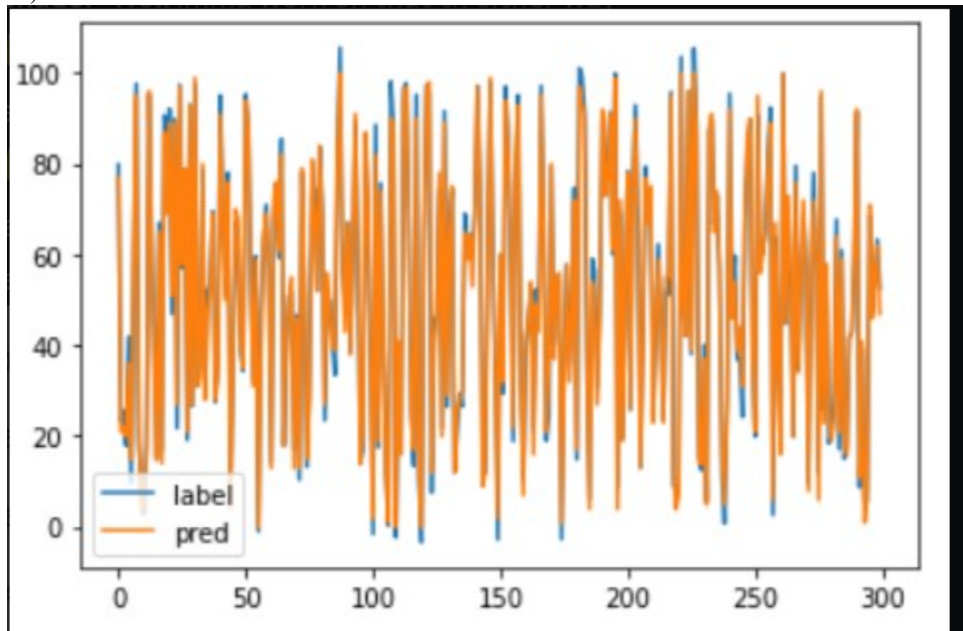
Joshua Clapp jclapp@tamu.edu

1a) data: first is train head n = 10 second is test n = 10

1b)

```
        x         y
0   24.0  21.549452
1   50.0  47.464463
2   15.0  17.218656
3   38.0  36.586398
4   87.0  87.288984
5   36.0  32.463875
6   12.0  10.780897
7   81.0  80.763399
8   25.0  24.612151
9    5.0   6.963319
        x         y
0   77  79.775152
1   21  23.177279
2   22  25.609262
3   20  17.857388
4   36  41.849864
5   15   9.805235
6   62  58.874659
7   95  97.617937
8   20  18.395127
9    5   8.746748
```

1c)

2a)

```
 2   -Andy Allanson,293,66,1,30,29,14,1,293,66,1,30,29,14,A,E,446,33,20,,A
 3   -Alan Ashby,315,81,7,24,38,39,14,3449,835,69,321,414,375,N,W,632,43,10,475,N
 4   -Alvin Davis,479,130,18,66,72,76,3,1624,457,63,224,266,263,A,W,880,82,14,480,A
 5   -Andre Dawson,496,141,20,65,78,37,11,5628,1575,225,828,838,354,N,E,200,11,3,500,N
 6   -Andres Galarraga,321,87,10,39,42,30,2,396,101,12,48,46,33,N,E,805,40,4,91.5,N
 7   -Alfredo Griffin,594,169,4,74,51,35,11,4408,1133,19,501,336,194,A,W,282,421,25,750,A
 8   -Al Newman,185,37,1,23,8,21,2,214,42,1,30,9,24,N,E,76,127,7,70,A
 9   -Argenis Salazar,298,73,0,24,24,7,3,509,108,0,41,37,12,A,W,121,283,9,100,A
10   -Andres Thomas,323,81,6,26,32,8,2,341,86,6,32,34,8,N,W,143,290,19,75,N
```

3a1)   The   linear   values   will   occasionally   repeat   positive   and   negative.

Coe [-8.98890267e-01 1.86488487e-01 7.69130419e-13 -1.80907871e-02
  4.30658981e-13 -4.31095009e-02 -5.00041306e-02 8.73701080e-03
  6.78644294e-02 7.91654030e-02 1.30108646e-01 1.21687095e-01
 -2.84480772e-13 4.68049655e-02 1.18583569e-01 -2.61721975e-01
  2.69062550e-13 2.92042778e-02 -4.86913597e-02 9.21115413e-02
 -4.73956539e-02 1.27336692e-01 -3.25683924e-12 1.08095254e-01
 -1.07239556e-03 -3.60605990e-12 2.57355248e-12 -9.99591371e-02
  2.66203726e-12 -4.86497867e-02 1.40522298e-01 1.78313597e-02
 -5.55111512e-17 -1.03609685e-01 1.36700827e-01 -5.80300115e-02
  1.35626831e-01 -7.76699249e-03 -1.03848861e-01 -8.60088736e-02
  8.54391219e-02 3.88578059e-16 1.24781938e-01 8.32667268e-17
 -1.66533454e-16 1.01915071e-01 1.02981076e-01 -8.81157809e-03
  1.15360162e-01 -2.15509913e-02 8.70187605e-01 -9.14398757e-01
 -5.22789546e-02 -5.54989486e-02 -4.12631188e-02 8.25194662e-02
  2.49800181e-16 1.33236979e-01 -9.85002214e-03 1.59184067e-01
  1.40691263e-01 -4.01685764e-02 -2.00064320e-02 -5.55111512e-17
  1.59836114e-01 1.62075128e-01 1.38777878e-16 -3.58156584e-02
 -5.28316572e-02 -2.77555756e-17 -3.61802216e-02 -3.34034860e-03
  5.92456645e-02 -3.41865627e-02 -1.46687749e-02 -5.55111512e-17
 -4.18326973e-02 1.12491475e-01 -1.11022302e-16 -5.61049568e-03
 -5.27702594e-02 1.39880192e-01 1.10961336e-02 1.04530292e-01
 -5.55372256e-02 -8.87847704e-01 -1.38777878e-16 -1.31838984e-16
 -1.30538556e-01 -5.29623732e-02 -2.09248356e-02 6.66783390e-02
  1.38313065e-01 5.69807535e-02 2.18334490e-01 -8.66910892e-01
 -1.66533454e-16 -2.33458278e-02 3.46944695e-18 -6.87851071e-02
 -6.79222435e-02 1.64273129e-01 1.02935152e-01 1.48375871e-01
  1.40672972e-01 1.55984935e-01 -1.24900090e-16 9.75751032e-01
  1.15166545e-01 9.44457394e-02 -8.15495837e-01 -4.87602223e-02
  1.94315585e-02 3.33066907e-16 1.11022302e-16 -5.88046210e-02
 -1.34940271e-01 0.00000000e+00 9.48134501e-02 0.00000000e+00
 -3.99958241e-02 1.38073738e-01 9.10243024e-01 1.03474672e-01
 -6.13269972e-02 1.16257972e-01 -5.22903356e-02 -1.01790687e-02
  4.01004682e-03 -6.31338618e-02 2.81025203e-16 5.38745133e-02
 -7.01510032e-02 8.59632921e-02 2.22044605e-16 1.97758476e-16
  1.14638940e-01 6.89214459e-02 8.94225172e-01 1.11022302e-16
  4.29164271e-02 -5.00862186e-02 -5.52247527e-02 1.06561734e-01
  1.17961196e-16 9.81107107e-02 -4.44089210e-16 -4.09726525e-02
  1.15862393e-01 9.58194312e-01 -1.06278432e-02 -5.60403280e-02
 -5.07533331e-02 -8.32667268e-17 6.57796530e-02 1.62224083e-01
  7.52507808e-02 -5.39580504e-02 -9.99353476e-02 1.12399708e-01
 -2.97875196e-02 -1.12397888e-01 1.32068638e-01 -2.83971469e-02
 -2.62257885e-02 1.02430735e-01 -1.82768543e-02 1.12890685e-01
 -1.94289029e-16 -8.58535722e-01 -2.05121635e-02 7.69287923e-02
  1.19581936e-01 -1.66533454e-16 8.32667268e-17 1.26752565e-01
 -5.98019787e-02 6.86595509e-02 5.55111512e-17 7.44811401e-02
 -9.83391610e-02 -6.52298180e-02 1.97170023e-01 1.16952568e-01
 -4.36060871e-02 -3.83485309e-02 -7.84697977e-02 -4.66836509e-02
 -1.18665770e-01 -9.31138006e-02 2.16274397e-01 8.73828710e-02
  1.06364754e-01 5.47388116e-02 -3.90594351e-02 4.33680869e-17
 -9.19186834e-01 -8.10020707e-03 -5.28652394e-02 -6.06808441e-02
  1.11022302e-16 -1.11022302e-16 -1.01092365e-02 0.00000000e+00
  1.31687721e-01 -3.14649557e-02 -7.08166458e-02 -1.65025700e-02
  0.00000000e+00 -1.54067330e-01 -7.63165006e-01 -3.10670953e-02
 -2.96083625e-02 6.70129290e-02 9.10352110e-02 0.00000000e+00
  0.00000000e+00 -3.14890127e-02 -2.80640605e-03 -1.45766513e-02
 -3.03401591e-02 0.00000000e+00 6.40328919e-02 -5.82525339e-02
 -5.41572935e-02 -5.75605604e-02 -7.13539658e-02 0.00000000e+00
 -1.15476450e-01 3.93115479e-02 1.45300069e-02 -8.63534624e-01
 -7.98957605e-02 5.67488649e-02 1.21293473e-01 8.69728519e-02
  0.00000000e+00 1.00516455e-02 0.00000000e+00 8.35977449e-02
  1.06417783e-01 0.00000000e+00 -2.89477071e-02 9.71617337e-02
 -5.62928006e-02 5.35575714e-02 -2.89387422e-02 -1.04920650e-01
  2.72861107e-02 5.25424179e-02 -6.27067191e-02 9.73210537e-02
  1.42857291e-02 1.12308464e-01 0.00000000e+00 0.00000000e+00
 -6.92201151e-02 -7.04690804e-02 9.61637969e-02 1.31817497e-02
  0.00000000e+00 -1.02826861e-01 -1.84472142e-01 -4.27224553e-01
  4.27224553e-01 9.39898225e-03 -9.39898225e-03 4.62088140e-04
 -1.17275709e-03 -1.08457179e-03 -6.37056483e-05 -7.60176767e-05
  9.80912524e-04 -3.51738505e-04 -1.33819332e-04 7.41394903e-04
  8.44763440e-04 -1.55932803e-04 -3.63137957e-04 -2.21745141e-04
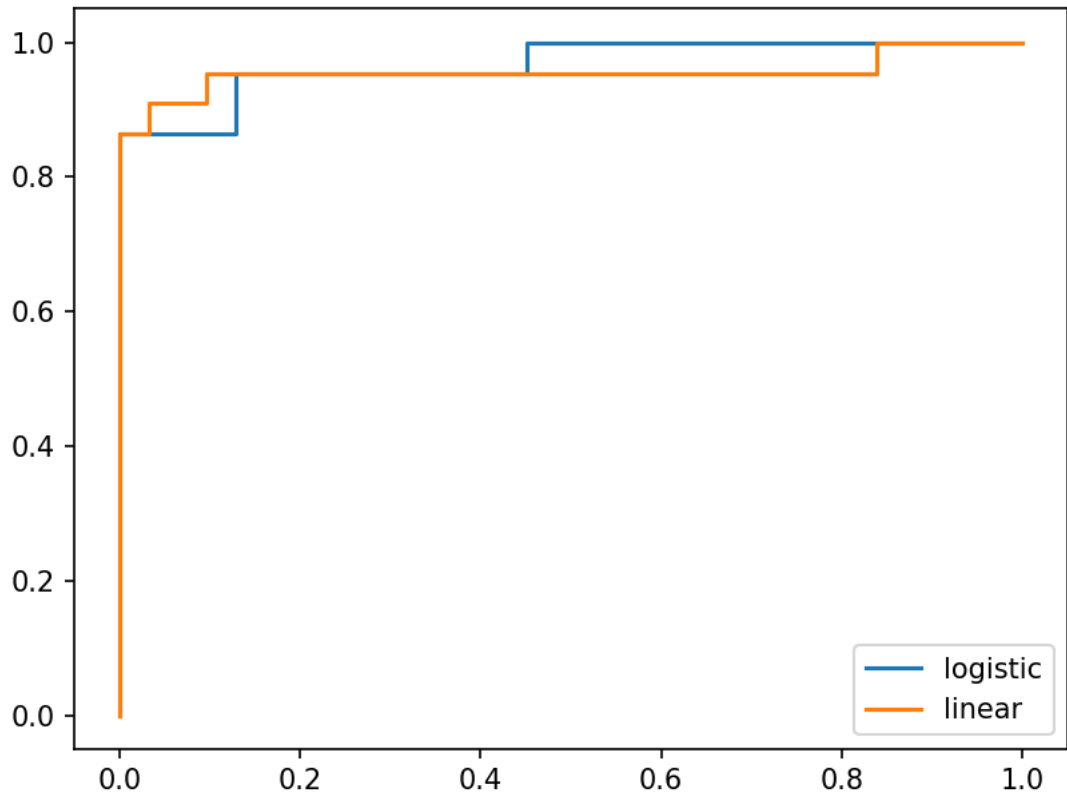 1.56603579e-05 6.15410726e-05 -2.25300948e-03 -4.59214382e-05] [[-7.
```

2

3a2) The logistical values here are all very similarly and have some repetitions.

1.56603579e-05  6.15410726e-05 -2.25300948e-03 -4.59214382e-05] [[-7.97852587e-01  2.25446693e-01  0.00000000e+00 -5.38627846e-02
 0.00000000e+00 -5.95106338e-02 -7.13405216e-02  5.34710347e-02
-1.71487231e-02  8.02838957e-02  1.12183522e-01  1.25474224e-01
 0.00000000e+00  6.84077117e-02  1.24635335e-01 -3.97521297e-01
 0.00000000e+00  2.63856156e-02 -5.75134997e-02  9.98177211e-02
-7.09786800e-02  1.55611595e-01  0.00000000e+00  1.06942608e-01
-4.08093348e-02  0.00000000e+00  0.00000000e+00 -8.96539478e-02
 0.00000000e+00 -4.82925275e-02  1.45440945e-01  3.16828994e-02
 0.00000000e+00 -8.50496442e-02  1.60580135e-01 -5.17699499e-02
 1.15708897e-01 -3.59359606e-02 -1.04503873e-01 -9.18734791e-02
 8.40649372e-02  0.00000000e+00  1.16635518e-01  0.00000000e+00
 0.00000000e+00  9.17424472e-02  9.62635912e-02 -5.24121277e-02
 1.37298717e-01 -5.37102902e-02  7.25648007e-01 -8.12784062e-01
-6.07458899e-02 -5.18168176e-02 -4.40550241e-02 -1.39263295e-02
 0.00000000e+00  1.56537871e-01 -4.82176835e-02  1.43707609e-01
 8.93852928e-02 -5.13093979e-02  3.71114210e-02  0.00000000e+00
 1.36132238e-01  2.23498302e-01  0.00000000e+00 -2.58339958e-02
-4.89958257e-02  0.00000000e+00 -5.79942241e-02 -3.04150884e-02
-2.41085691e-02 -4.38844486e-02 -3.92000386e-02  0.00000000e+00
-7.43245637e-02  1.36012497e-01  0.00000000e+00  2.22398948e-02
-6.40100449e-02  1.47430852e-01 -2.79235767e-02  8.32178215e-02
-5.28626085e-02 -7.50809121e-01  0.00000000e+00  0.00000000e+00
-1.78475340e-01 -6.62989221e-02 -5.55754445e-02  8.74683108e-02
 1.18402462e-01  7.78533091e-02  2.69090210e-01 -7.49531500e-01
 0.00000000e+00 -5.33615084e-02  0.00000000e+00 -7.10020683e-02
-6.78002500e-02  2.00270940e-01  9.46908483e-02  1.41014309e-01
 1.22643864e-01  1.56210181e-01  0.00000000e+00  8.54539444e-01
 1.28931112e-01  8.15318061e-02 -6.49915624e-01 -3.79651948e-02
-2.40286749e-02  0.00000000e+00  0.00000000e+00 -5.38023578e-02
-1.24230515e-01  0.00000000e+00  1.01575870e-01  0.00000000e+00
-5.67400547e-02  1.19376466e-01  8.07229514e-01  8.55908187e-02
-5.73715271e-02  1.12416415e-01 -4.64140113e-02 -4.46024202e-02
 3.68554539e-02 -5.53525218e-02  0.00000000e+00  5.70972505e-02
-6.72182897e-02  1.02370085e-01  0.00000000e+00  0.00000000e+00
 9.11683100e-02  9.25018084e-02  7.69914830e-01  0.00000000e+00
 4.51455225e-02 -4.47929318e-02 -6.63172281e-02  1.31186794e-01
 0.00000000e+00  1.02988192e-01  0.00000000e+00 -4.20547052e-02
 9.11644187e-02  8.67088770e-01 -1.14243653e-01 -5.62914253e-02
-5.41019335e-02  0.00000000e+00  8.31534642e-02  1.76241756e-01
 7.73484222e-02 -5.00429608e-02 -9.54935533e-02  9.54924417e-02
-5.84409393e-02 -9.11098643e-02  1.09547273e-01 -3.73218568e-02
-6.00364865e-02  8.09945035e-02 -3.46281304e-02  1.05092534e-01
 0.00000000e+00 -7.16828840e-01 -5.69580226e-02  7.67833163e-02
 1.15188006e-01  0.00000000e+00  0.00000000e+00  9.99116707e-02
-5.90818916e-02  7.37477803e-02  0.00000000e+00  7.51071472e-02
-7.48210218e-02 -7.35247301e-02  1.85676002e-01  1.19217953e-01
-5.68436159e-02 -5.08304103e-02 -9.85142003e-02 -5.05236528e-02
 1.91914723e-03 -8.55008756e-02  2.33198632e-01  9.89817838e-02
 8.17269655e-02  5.99893603e-02 -4.60480894e-02  0.00000000e+00
-8.12753771e-01 -2.87643551e-02 -6.94848944e-02 -1.00551513e-01
 0.00000000e+00  0.00000000e+00 -3.73154963e-02  0.00000000e+00
 1.20827191e-01 -4.46113505e-02 -5.52971452e-02 -4.68423360e-02
 0.00000000e+00 -1.74069572e-01 -5.44237750e-02 -5.01332295e-02
-4.21898612e-02  5.81232319e-02  9.85884820e-02  0.00000000e+00
 0.00000000e+00 -5.22374744e-02  4.23660631e-02 -4.98560355e-02
-4.32606566e-02  0.00000000e+00  8.55310397e-02 -6.68404168e-02
-5.54238001e-02 -6.29294484e-02  1.51456707e-02  0.00000000e+00
-1.12336249e-01  5.15709905e-02  9.78033971e-02 -7.34520062e-01
-7.04996991e-01  7.12883282e-02  7.10494891e-02  8.81746457e-02
 0.00000000e+00 -3.67804592e-02  0.00000000e+00  8.65066311e-02
 1.18887544e-01  0.00000000e+00 -4.34390669e-02  9.08728380e-02
-6.10795429e-02  5.98816000e-02 -4.22296716e-02 -7.86975541e-02
 5.24223909e-02  5.50430767e-02 -6.74390520e-02  9.77036766e-02
 5.56059636e-02  1.27304126e-01  0.00000000e+00  0.00000000e+00
-7.01575768e-02 -7.15406652e-02  8.98060046e-02  4.23439168e-02
 0.00000000e+00 -8.68373549e-02 -2.23573517e-01 -2.50877041e+00
 2.33275075e+00 -2.95770869e-02 -1.46442572e-01  4.94461509e-03
-1.48329571e-02 -1.31671664e-02 -3.87094626e-03 -6.81914413e-04
 1.17154229e-02 -5.43728067e-02 -1.64463660e-03  1.02933858e-02
 1.03233229e-02 -3.02528000e-03 -5.46632192e-03 -1.79415451e-03
 2.70547452e-04  7.07202978e-04 -1.89276702e-02 -4.35951083e-04]]

3a) The linear and logistical regression values are not the same and each has some similarities within some of them. The linear values seem to repeat a positive then a negative and the logistical has a clear pattern.



3b)

3c) Optimal threshold linear = 0.11195390516759324 logistical = 0.7032913459513034. I calculated the optimal threshold that takes the tpr and fpr, finds the difference and applies it to the absolute maximum.

4a) The features do change each fold. This is from the stratified split.

4b) auc score linear mean: 0.9316190476190476 0.8897016486381549 mean+standard error: 0.9735364465999402
auc score log mean: 0.9316190476190476 mean-standard error: 0.8897016486381549 mean+standard error: 0.9735364465999402

4c) auc score linear f1 mean: 0.9276174299507037 mean-standard error: 0.8838143294210034 mean+standard error: 0.9714205304804041
auc score log f1 mean: 0.9109418028080798 mean-standard error: 0.8708057637611156 mean+standard error: 0.951077841855044

4