1.

Chart

Description automatically generated

Number of positive points: 958

Number of negatives points: 1042

Final correctness: 2000

Total iteration: 31

Final w: [29. 1.92092337 53.66631267]

Liner regression coefficients: [ 0.24591466 -0.00937018 0.0790168 ]

From the graphs it can be inferred that PLA and linear regression achieves very close solution. Consider a hyperplane y = w0 + w1x1 + w2x2 for the -1 and 1 points given x in the space. Once we find the w through linear regression, we let w0 + w1x1 + w2x2 = 0, this hyperplane should separate the points approximately.

2.

Chart, line chart

Description automatically generated

Final correctness: 2000 . Total iteration: 18  
Final w: [ 6. -0.89097195 47.94025173]  
Final correctness: 2000 . Total iteration: 41  
Final w: [15. -0.63429873 82.0719898 ]  
Final correctness: 2000 . Total iteration: 28  
Final w: [10. -1.72886399 64.03266339]  
Final correctness: 2000 . Total iteration: 27  
Final w: [11. -1.77381982 62.10530211]  
Final correctness: 2000 . Total iteration: 27  
Final w: [11. -1.77381982 63.70530211]  
Final correctness: 2000 . Total iteration: 26  
Final w: [12. -1.81877565 61.17794083]  
Final correctness: 2000 . Total iteration: 29  
Final w: [13. -0.60062561 68.41241207]  
Final correctness: 2000 . Total iteration: 20  
Final w: [ 8. -0.94113792 55.47314254]  
Final correctness: 2000 . Total iteration: 20  
Final w: [ 8. -0.94113792 56.67314254]  
Final correctness: 2000 . Total iteration: 22  
Final w: [10. -0.73920639 59.74861391]  
Final correctness: 2000 . Total iteration: 16  
Final w: [ 6. -1.10571191 53.72224028]  
Final correctness: 2000 . Total iteration: 19  
Final w: [ 9. -1.26842599 52.63027033]  
Final correctness: 2000 . Total iteration: 19  
Final w: [ 9. -1.26842599 53.63027033]  
Final correctness: 2000 . Total iteration: 19  
Final w: [ 9. -1.26842599 54.63027033]  
Final correctness: 2000 . Total iteration: 16  
Final w: [ 6. 0.13843134 55.84381562]  
Final correctness: 2000 . Total iteration: 13  
Final w: [ 5. -0.39078826 46.52547204]  
Final correctness: 2000 . Total iteration: 13  
Final w: [ 5. -0.39078826 47.32547204]  
Final correctness: 2000 . Total iteration: 23  
Final w: [13. 0.87921332 61.61639105]  
Final correctness: 2000 . Total iteration: 12  
Final w: [ 6. -0.45419319 43.89948258]  
Final correctness: 2000 . Total iteration: 12  
Final w: [ 6. -0.45419319 44.49948258]  
Final correctness: 2000 . Total iteration: 12  
Final w: [ 6. -0.45419319 45.09948258]  
Final correctness: 2000 . Total iteration: 12  
Final w: [ 6. -0.45419319 45.69948258]  
Final correctness: 2000 . Total iteration: 12  
Final w: [ 6. -0.45419319 46.29948258]  
Final correctness: 2000 . Total iteration: 12  
Final w: [ 6. -0.45419319 46.89948258]  
Final correctness: 2000 . Total iteration: 11  
Final w: [ 7. -0.51759811 41.27349311]

It generally takes more iterations for PLA to converge when *sep* is small. Looks like ||w|| is increasing when *sep* is decreasing. The R term is fixed due to x are fixed. The p is the minimum of ynwtxn, which is less affected by the change of ||w||. So, the overall effect is to increase the time to converge for PLA.

3.

Text, letter

Description automatically generated

Text, letter

Description automatically generated

Text, letter

Description automatically generated

A piece of paper with writing

Description automatically generated with medium confidence

Text, letter

Description automatically generated

Text

Description automatically generated

e.) Calculations using the program:

E(u+du,v+dv) with (du,dv) from E\_1 approx. = 2.2508597349929693

Newton direction: (0.4587778126549621, 0.19880371881715023)

E(u+du,v+dv) = 1.8904907903020918  
  
Optimal direction by minimizing E(u+du,v+dv): (0.4355689881974021, 0.2455191571358361)

Minimal E(u+du, v+dv): 1.8684370301391746

4.

5.

Consider a given H

If the best approximation from H is less complex than the initial target function, then when we increase the complexity of f, the deterministic noise in general should increase, since it'll be harder for functions in H to fit the target function. There'll be a higher tendency to overfit.

If the best approximation from H is more complex than the initial target function, then when we increase the complexity of f, the deterministic noise in general may decrease first, reducing the deterministic noise and there'll be a lower tendency to overfit. But once the complexity of f exceeds the best function approximation from H, and if we continue increase the complexity of f, we will increase the deterministic noise and thus increase the tendency to overfit.

(b) Given a fixed f

If the best approximation from H is less complex than the target function, then when we decrease the complexity of H, we increase the deterministic noise thus increasing the tendency of overfit.

If the best approximation from H is more complex than the target function, then when we decrease the complexity of H, we will decrease the deterministic noise thus decreasing the tendency of overfit. Well, if we continue to decrease the complexity of H, passing the point where its complexity is equal to f, we start to increase the deterministic noise again and thus increasing overfit.

6.

Text, letter

Description automatically generatedA piece of paper with writing

Description automatically generated with medium confidenceA piece of paper with writing

Description automatically generated with medium confidence