Kim, Joo Eun (June)
12/7/18

CS 156a Final

7. The attached code generated E_{in} values of 0.07625840076807022, 0.09107118365107666, 0.08846523110684405, 0.07433822520916199, and 0.08832807570977919 respectively for the classifiers in [a] through [e]. The lowest value, 0.07433822520916199, corresponds to [d] 8 versus all.

Answer: [d]

8. The attached code generated E_{out} values of 0.10662680617837568, 0.02192326856003986, 0.09865470852017937, 0.08271051320378675, and 0.09965122072745392 respectively for the classifiers in [a] through [e]. The lowest value, 0.02192326856003986, corresponds to [b] 1 versus all.

Answer: [b]

9. The attached code generated the following results:

0 versus all without transform
E_in: 0.10931285146070498 E_out: 0.11509715994020926

0 versus all with transform
E_in: 0.10231792621039638 E_out: 0.10662680617837568

1 versus all without transform
E_in: 0.01522424907420107 E_out: 0.02242152466367713

1 versus all with transform
E_in: 0.012343985735838706 E_out: 0.02192326856003986

2 versus all without transform
E_in: 0.10026059525442327 E_out: 0.09865470852017937

2 versus all with transform
E_in: 0.10026059525442327 E_out: 0.09865470852017937

3 versus all without transform
E_in: 0.09024825126868742 E_out: 0.08271051320378675

3 versus all with transform
E_in: 0.09024825126868742 E_out: 0.08271051320378675

4 versus all without transform
E_in: 0.08942531888629818 E_out: 0.09965122072745392

4 versus all with transform
E_in: 0.08942531888629818 E_out: 0.09965122072745392

5 versus all without transform
E_in: 0.07625840076807022 E_out: 0.07972097658196313

5 versus all with transform

E_in: 0.07625840076807022 E_out: 0.07922272047832586

6 versus all without transform
E_in: 0.09107118365107666 E_out: 0.08470353761833582

6 versus all with transform
E_in: 0.09107118365107666 E_out: 0.08470353761833582

7 versus all without transform
E_in: 0.08846523110684405 E_out: 0.07324364723467862

7 versus all with transform
E_in: 0.08846523110684405 E_out: 0.07324364723467862

8 versus all without transform
E_in: 0.07433822520916199 E_out: 0.08271051320378675

8 versus all with transform
E_in: 0.07433822520916199 E_out: 0.08271051320378675

9 versus all without transform
E_in: 0.08832807570977919 E_out: 0.08819133034379671

9 versus all with transform
E_in: 0.08832807570977919 E_out: 0.08819133034379671

From this we can see that [a] through [d] are not true, and [e] is true, as the transform improves the out-of-sample performance of '5 versus all' by 0.625%.
Answer: [e]

10. The attached code generated the following results:

lamda: 0.01
E_in: 0.004484304932735426 E_out: 0.02830188679245283

lamda: 1
E_in: 0.005124919923126201 E_out: 0.025943396226415096

We can see that E_{in} goes down and E_{out} goes up from $\lambda = 1$ to $\lambda = 0.01$.
Answer: [a]

11. Plotting the transformed points $\mathbf{z}_n$ in the Z space:

Kim, Joo Eun (June)
12/7/18

12. The attached code uses sklearn.svm package based on quadratic programming, which produces the following values representing $y_i\alpha_i$: [[-1.00000000e+07 -8.88888896e+06 1.00000000e+07 8.88888894e+06 2.22162241e-02]]. Since there are 5 nonzero values of α_i, there are 5 corresponding x_n values, which are the support vectors.

Answer: [c]

13. The attached code generated 0% for 1000 runs.

Answer: [a]

14. The attached code generated 84%.

Answer: [e]

15. The attached code generated 75%.

Answer: [d]

16. The attached code generated the following frequency distribution for 1000 runs:

{'a': 156, 'b': 118, 'c': 92, 'd': 456, 'e': 3}

Answer: [d]

17. The attached code generated the following frequency distribution for 1000 runs:

{'a': 181, 'b': 155, 'c': 304, 'd': 190, 'e': 2}

Answer: [c]

18. The attached code generated 1%.

Answer: [a]

19. Given a prior P(h = f), we can get the posterior P(h = f | D) ∝ P(D | h = f) P(h = f). Since we have the dataset D (one person chosen from the population who had a heart attack), we can calculate a specific value of P(D | h = f) for a hypothesis h, which is the probability of collecting a dataset like that (where the one person chosen had a heart attack) given the hypothesis truly reflected the probability of getting a heart attack for people in a certain population. Since our only data point had a heart attack, that probability P(D | h = f) would increase linearly over h ∈ [0, 1]; there is 0 probability of getting such a data point--a person who had a heart attack--if the true probability of getting a heart attack was 0, 1 probability of finding a person who had a heart attack if the true probability of getting a heart attack was 1, and the probability of finding a person who had a heart attack would linearly increase with the true probability of getting a heart attack, reflected by h, linearly increasing. And with our assumption that P(h = f) is uniform over h ∈ [0, 1], the posterior P(h = f | D) ∝ P(D | h = f) P(h = f) linearly increases over [0, 1], since P(D | h = f) linearly increases over [0, 1] and P(h = f) is uniform over [0, 1] (think of multiplying a linearly increasing function with a constant function, which gives a linearly increasing function).

Answer: [b]