Jacob Snyder

CS156a Final

1. E
   For degree 1 we have 2 terms: (x_1, x_2)
   For degree 2 we have 5 terms: $(x_1, x_2, x_1x_2, x_1^2, x_2^2)$
   For degree 3 we have 9 terms: $(x_1, x_2, x_1x_2, x_1^2, x_2^2, x_1^2x_2, x_1x_2^2, x_1^3, x_2^3)$
   Therefore, following the pattern we should have $2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 + 10 + 11 = 65$ terms for $Q = 10$. Therefore the answer is E.

2. D
   The logistic regression model being used could result in gbar not being in H because of the odd properties of the sigmoid function. The average of several hypotheses in H could result in an average hypothesis that doesn't match the sigmoid function.

3. D
   D is false because there can be other reasons that $E_{out} - E_{in}$ is large; for example the possibility that the training set is a poor representative sample but the correct amount of fit is being used.

4. D
   Stochastic noise does not depend on the hypothesis set, by definition. This can be found in the lecture slides (lecture 11, slide 16).

5. A
   By definition, $\boldsymbol{w}_{lin}$ is a solution to the minimization of $\frac{1}{N}\sum_{n=1}^{N}(\boldsymbol{w}^T\boldsymbol{x}_n - y_n)^2$. Therefore, if we add the constraint that $\boldsymbol{w}_{lin}^T\Gamma^{\mathrm{T}}\Gamma\mathbf{w}_{\mathrm{lin}} \leq C$, $\boldsymbol{w}_{lin}$ also satisfies the constraints of regularization. Therefore, it is true that $\boldsymbol{w}_{lin} = \boldsymbol{w}_{reg}$ so the answer is A.

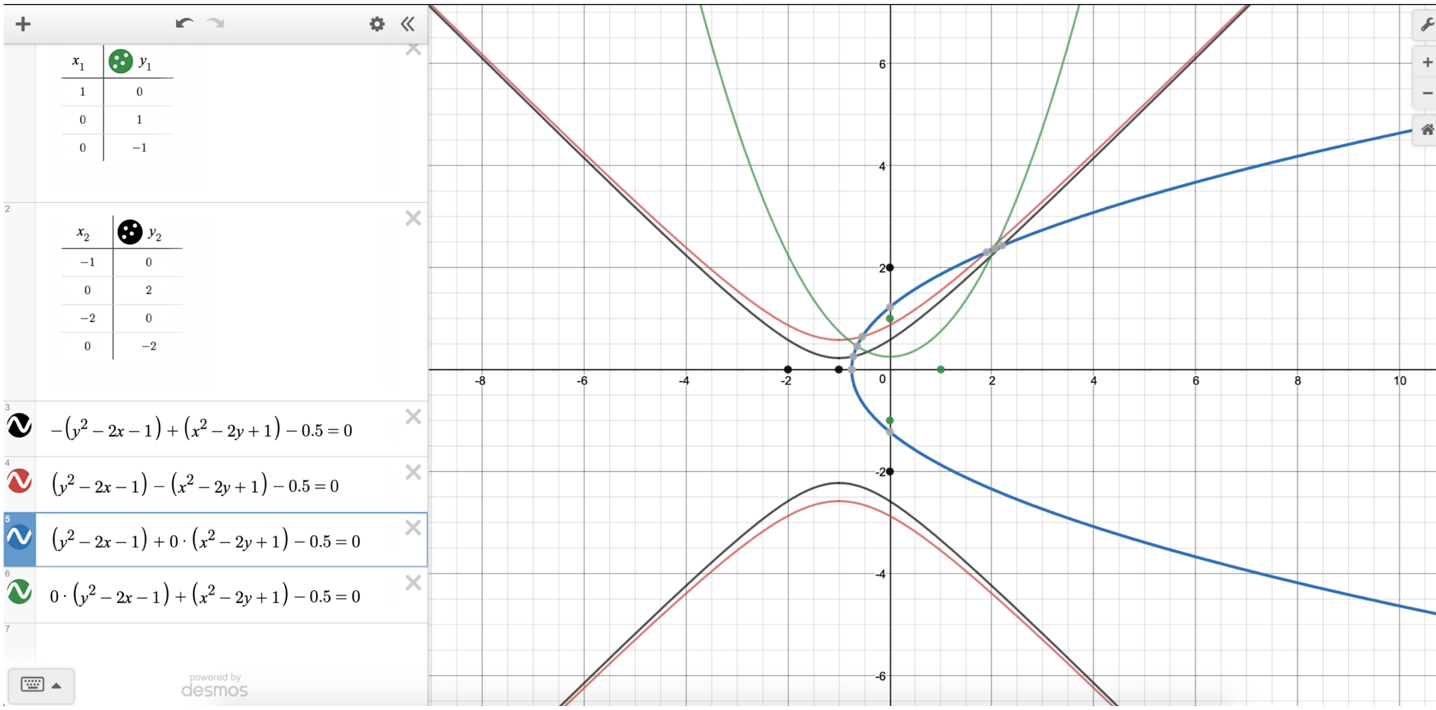
6. B
   Soft-order constraints are translated into augmented error. This can be found in the slides (lecture 12). This is what generates the term $\frac{\lambda}{N}\boldsymbol{w}^T\boldsymbol{w}$ in the error function for regularization.

7. D
   Justified in Jupyter Notebook

8. B
   Justified in Jupyter Notebook

9. E
   Justified in Jupyter Notebook

10. A
    Justified in Jupyter Notebook
11. C
    Using desmos to plot everything we have:



The only plot that correctly separates the points is the blue one, which corresponds to the weights from C. Therefore the answer is C.

12. C
Justification in Jupyter Notebook

13. A
Justified in Jupyter Notebook

14. E
Justified in Jupyter Notebook

15. D
Justified in Jupyter Notebook

16. D
Justified in Jupyter Notebook

17. C
Justified in Jupyter Notebook

18. A
Justified in Jupyter Notebook

19. B
We know that $P(h = f|D) \propto P(D|h = f) * P(h = f)$.
Given that $P(h = f)$ is constant and $P(D|h = f)$ is linear (because it's just h by definition), we have that $P(h = f|D)$ must be linear, so the answer is B.

20. C
Because we're using mean squared error, the error of the average hypothesis will always be closer to the error of the better hypothesis than to the error of the worse hypothesis (this is intuitive; without squaring the gap between average hypothesis and worse hypothesis is the same as between average hypothesis and good hypothesis by nature of the average, but once you square in order to get error, the larger error

value for the worse hypothesis will cause it to grow much faster than the average hypothesis or the good hypothesis; see an exponential curve). Therefore we have that the mean squared error of the average hypothesis can't be worse than the average of the error of the good and bad hypotheses.