FINAL EXAM

Problem 1

e is the correct answer.
We can start by measuring the dimensionality and then showing that it surpasses 100, thus not being an answer choice. Let the coordinates in our original space be (x_1, x_2). In the Z space, we will have 11 coordinates of form x_1^n where n goes from 1 to 10 and likewise, 11 coordinates of form x_2^n where n goes from 1 to 10. There will be 11 coordinates of form $\mathrm{x}_1^n\mathrm{x}_2^n$ where n goes from 1 to 10. There are 9 coordinates of form $\mathrm{x}_1^n\mathrm{x}_2$ where n goes from 2 to 10, 8 coordinates of form $\mathrm{x}_1^n\mathrm{x}_2^2$ where n goes from 3 to 10, and similarly, 7, 6, 5, 4, 3, 2, and ultimately 1 coordinate of form $\mathrm{x}_1^{10}\mathrm{x}_2^9$. Thus, there are 9 + 8 + 7 + 6 + 5 + 4 + 3 + 2 + 1 = 45 coordinates where x_1 has a higher exponent than x_2 and likewise, 45 coordinates where x_2 has a higher exponent than x_1. Summing all these together, we have 45 + 45 + 11 + 11 + 11 which is clearly greater than 100 and not one of the answer choices. Thus, we get choice e.

Problem 2

c or d is the correct answer.
By the process of elimination, I know that a and b are not valid because weighted sum of hypothesis from a singleton H will always return that one hypothesis from H. Likewise, for constant, real-valued function, we have the same case (averaging over constant hypothesis will return another constant hypothesis). Thus, I am left with choice c or d.

Problem 3

d is the correct answer.
Per slide 12 of lecture 11 on overfitting, it claims the overfit measure to be: $\mathrm{E}_{\mathrm{out}}(\mathrm{g}_1)$ - $\mathrm{E}_{\mathrm{out}}(\mathrm{g}_2)$ where g_1 and g_2 are two hypothesis. This contrasts answer choice d.

Problem 4

d is the correct answer.
Per slide 16 of lecture 11, it states that one of the main differences between deterministic noise and stochastic noise is that deterministic noise depends on H. This implies that stochastic noise does not depend on the hypothesis set, answer choice d.

Problem 5

a is the correct answer.
As shown geometrically in slide 9 of lecture 12, if $\mathrm{w}_{\mathrm{lin}}$ lies within the constraint ($\leq$ C), then we simply pick $\mathrm{w}_{\mathrm{reg}}$ to be equal to $\mathrm{w}_{\mathrm{lin}}$ because that's when there is minimum E_{in}.

Problem 6

b is the correct answer.
From slide 10 of lecture 12, we see a parallel between regularization of polynomial models and augmented error. Minimizing $\mathrm{E}_{\mathrm{aug}}(\mathrm{w})$ which is $\mathrm{E}_{\mathrm{in}}(\mathrm{w}) + \frac{\lambda}{N}\mathrm{w}^T\mathrm{w}$ solves the problem of minimizing E_{in} (w) subject to a constraint ($\mathrm{w}^T\mathrm{w} \leq \mathrm{C}$). Thus, we get answer choice b.

Problem 7

d is the correct answer.
See attached code. I coded up linear regression with regularization, choosing $\lambda = 1$ and found that the the lowest E_{in} came from 8 versus all, an error of 0.07433822520916199.

Problem 8

b is the correct answer.
See attached code. Using the same code as problem 7, but applying a transformation and then getting the lowest $\mathrm{E}_{\mathrm{out}}$ resulted in 1 versus all having the lowest error of 0.02192326856003986.

Problem 9

e is the correct answer.
See attached code. Calculating the $\mathrm{E}_{\mathrm{out}}$ for '5 versus all' with and without transformation resulted in values of 0.07972097658196313 and 0.07922272047832586, respectively. Thus, the transformation did lower $\mathrm{E}_{\mathrm{out}}$, but by certainly less than 5%.

Problem 10

a is the correct answer.
See attached code. For $\lambda = 1$, $\mathrm{E}_{\mathrm{out}} = 0.025943396226415096$ and $\mathrm{E}_{\mathrm{in}} = 0.005124919923126201$. For $\lambda = 0.01$, $\mathrm{E}_{\mathrm{out}} = 0.02830188679245283$ and $\mathrm{E}_{\mathrm{in}} = 0.004484304932735426$. By process of elimination or noticing that $\mathrm{E}_{\mathrm{out}}$ increases from $\lambda = 1$ to $\lambda = 0.01$, we get choice a.

Problem 11

c is the correct answer.
After transformation, we get the points (-3,2), (0, -1), (0, 3), (1, 2), (3,-3), and (3,5). I plotted these points by hand and in mathematica and based on the values generated from their target function, saw that by simple geometric analysis, choice c had the values maximizing the margin. We also notice that the other choices do not separate the data properly, leaving us with choice c.

Problem 12

c is the correct answer.
See attached code. After running the code, the output was 5, corresponding to answer choice c.

Problem 13

a is the correct answer.
See attached code. After running the code (1000 runs), I found that the number of times E_{in} = 0 to be 0, which is less than 5% of the time.

Problem 14

e is the correct answer.
See attached code. After setting up the regular RBF and the SVM with hard-margins, I compared out of sample performance on 500 randomly generated "out of sample" points, and found that when k = 9, the kernel form beats the regular form 85.6% of the time, or choice e.

Problem 15

d is the correct answer.
See attached code. Using the same code as problem 14, but changing from k = 9 to k = 12 resulted in 80% of the time where the kernel form beat the regular form, or choice d.

Problem 16

d is the correct answer.
See attached code. Over 100 runs, I kept track of how many times E_{in} and $\mathrm{E}_{\mathrm{out}}$ decreased when I went from k = 9 to k = 12 and found that these events happened 71 and 82 times, respectively. This means that for the most part, both E_{in} and $\mathrm{E}_{\mathrm{out}}$ went down, or choice d.

Problem 17

c is the correct answer.
See attached code. Over 100 runs, I kept track of how many times E_{in} and $\mathrm{E}_{\mathrm{out}}$ decreased when I went from $\gamma = 1.5$ to $\gamma = 2$ and found that these events happened 30 and 42 times, respectively. This means that the number of times E_{in} and $\mathrm{E}_{\mathrm{out}}$ increased was 70 and 58 times, respectivly. Thus, for the most part, both E_{in} and $\mathrm{E}_{\mathrm{out}}$ went up, or choice c.

Problem 18

a is the correct answer.
See attached code. Over 500 runs, I kept track of how many times E_{in} for regular RBF with k = 9 and $\gamma = 1.5$ was 0 and found that such occurred 3.4% of the runs, which is $\leq$ 10%, or choice a.

Problem 19

b is the correct answer.
We know that the posterior $\propto$ likelihood x prior. We know that from the problem, our prior is uniform, but our likelihood increases linearly over [0,1] because the probability of our data (one example of heart attack) given h = f increases linearly as h = f gets closer to 1, meaning 100% probability of getting a heart attack.

Problem 20

c is the correct answer.
We are forming an aggregate hypothesis where each of the two hypothesis (g_1 and g_2) contribute equally. Thus, the E_{out} of our new hypothesis cannot be worse than the average of the E_{out} values for g_1 and g_2.