FINAL: 90 Minutes

Last Name: _____

First Name: _____

RIN: _____

Section: 4100 / 6100 (circle one)

Answer **ALL** questions.
NO COLLABORATION or electronic devices. Any violations result in an F.
NO questions allowed during the test. Interpret and do the best you can.
ALWAYS show your work and justify each answer.

GOOD LUCK!

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	TOTAL

- **You do not have time to waffle. So, don't waffle.**
- **Keep your answers precise and concise.**
- **Each question is worth 1 point.**

1. What is the two step approach to learning and why do we do it that way?

2. You're a Facebook Troll. All your FB-friends have exciting lives based on their posts. In comparison, your life seems mundane and you get depressed. What learning from data trap could you be in?

3. Why is it a good idea to preprocess the data with input normalization?

4. Derive the growth function for the 1-dimensional positive intervals. Explain how to get the VC-dimension, and give the VC-dimension.

5. Based on your VC-dimension for positive intervals, what is the error bar linking E_{in} to E_{out}?

6. You decided to use 8th order polynomials to fit the data, to give your learning lots of flexibility. You suspect you might overfit the data. What does that mean?

7. In the previous problem, what could you do to help the learning in case of overfitting?

8. What is a validation set. Why do we use it?

9. What are the tradeoffs in choosing the validation set? Why should it be small? Why large?

10. Define the leave one out cross-validation error, E_{CV}?

11. Prove that E_{CV} based on N data points is an unbiased estimate of your expected out-of-sample error when you learn on $N-1$ data points.?

12. The nearest neighbor algorithm was at most a factor of 2 away from optimal. Let E^*_{out} be the optimal probabilty of error at test point $\mathbf{x}$. Prove that the 3-nearest neighbor algorithm is near-optimal. That is, for N large enough, the out of sample error at $\mathbf{x}$ is bounded by

$$E_{\text{out}}(\text{3-NN}) \leq E^*_{\text{out}}(1 + 3E^*_{\text{out}}).$$

If optimal performance is 1% error, how bad can 3-NN be? (In your proof, state any assumptions.)

13. Explain why the 1-hidden layer neural network is more powerful than the K-RBF network. What are the pros and cons of this power?

14. Explain why the optimal hyperplane with maximum margin performs well even in very high dimensional spaces where a random perceptron won't.

15. Why is it computationally tractable to run the optimal hyperplane using a feature transform into very high, even infinite, dimensions.

SCRATCH