Final Prep

Format:

1. Your final exam will be an in-person 1-1 oral exam. There are two main reasons:
   (a) The oral exam will give you practice for technical interviews.
   (b) Studies show that students get better grades on oral exams and retain knowledge longer.
2. The exam will be worth 30 points (and replace your lowest midterm grade).
3. My goal is for each exam to take about 30 minutes (but you are welcome to use the entire scheduled 60 minute block if you'd like).
4. You may not reference any notes during the exam.

The Questions:

1. There will be 4 questions from previous midterms:
   (a) 2 questions from the T/F/O section of midterm 1 and 2 questions from midterm 2.
   (b) They will be randomly selected by rolling a 10-sided dice.
   (c) Each worth 5 points
      i. For full credit, you will have to state the correct answer to each question and *explain why that answer is correct.*
      ii. The explanation must explicitly define all of the terms you use that you were required to memorize for a quiz.
      iii. I may ask short follow-up questions to ensure you understand the material. For example,
         A. On midterm 1, if the question is based on the $\mathbf{P}$ matrix, I might ask you "would the answer still be the same if we used $\bar{\mathbf{P}}$ instead?" Or if the question uses the word "irreducible" I might ask you if it is still true if we change that to "primitive."
         B. On midterm 2, I might ask you to change the hypothesis class of a question or change the kernel of a question.
2. There will be one open-ended interview style question.
   (a) It is worth 10 points.
   (b) Everyone will have the same question:

   Question. Our company is building a content moderation system for a new social media platform. In order to do this, we are building a text classification model that outputs whether the input text is abusive or not. For simplicity, we plan on using a logistic regression model, but we're not yet sure what types of features we should use. We are considering two options: 1-hot encoding with feature hashing (with vowpal wabbit) or word2vec (with scikit-learn). Which would you choose and why?

   Your answer must include:

      i. a drawing of the model complexity curve
      ii. a discussion about where the two models in the question fall on the model complexity curve

   (c) I will ask 1-2 followup questions that will depend on how you answer the previous question. Potential followup questions include:
      i. You mentioned VC dimension... what is that?
      ii. Explain which kernels you might want to use on this problem and why.
      iii. Are there other models you might recommend we use?
      iv. What would be the advantages/disadvantages of switching to a different model, like XXXX?

v. Explain how you would evaluate your model. (For example, using a training/validation set.)
vi. Define the different types of regularization and explain how they relate to your discussion. (My expectation is that you would be able to define the soft order constraint and augmented error regularization methods.)
vii. Explain how a pagerank feature could be used in addition to the text feature to make the model more accurate. Would this change your choice of model?

Tips:

1. You know exactly what the first question will be (the open-ended question above).
   (a) Prepare in-advance what you want to say and write
   (b) Practice saying and writing your response in front of an audience.
   (c) This will ensure that you're off to a good start, and build your confidence for the remaining questions.
2. Find excuses to pause and think
   (a) Read the question out loud
   (b) As you read, write down key information on the whiteboard
   (c) Find a keyword in the prompt that you know, and say something about that
   (d) Set down your whiteboard marker and pick up a different one
   (e) Bring a water bottle, and take a sip when you need to think
   (f) It's okay to say "Hmmm.... let me think about that for a second..."
      i. say is a bit slowly for extra time
      ii. don't stay silent for more than about 5 seconds or so before starting to think out loud
3. Boardsmanship
   (a) Not explicitly graded... but a good presentation will make it easier for me to find reasons to give you points.
   (b) Write something for every problem
   (c) Write slowly and clearly, start in the upper left hand corner of the board
   (d) Capital letters should be $\sim$ 3in tall; diagrams should be large
      i. assume the interviewer has bad eyesight
   (e) Use colors to differentiate parts of your answer. There are many ways to do that, but some ideas are:
      i. Use 1 color for the information from the problem, different color for your solution
      ii. Use multiple colors in any figures you draw
      iii. If you have multiple definitions, use multiple a different color for each definition
   (f) Between problems, totally erase the whiteboard
   (g) (Ideally) talk while you write, and position your body so that the audience can hear you and see what you're writing
      i. also okay to say "I'm going to write for a bit and then explain what I've written"
      ii. not good to just start writing without saying anything
4. Content:
   (a) Ensure you have memorized the definitions from the quizzes
5. Wednesday we will review the midterm3, but also have time for students to practice giving responses. These won't be recorded.