Final, CS-GY 6923

Basic Info

Date: Thursday 5/12 from 11am - 12:30pm **Location:** Our usual classroom **Supplies needed:** Pencil or pen. I will provide scrap paper. **Supplies allowed:** One-double sided sheet of paper containing whatever you like. Equations, notes, examples, etc. This can be handwritten or typewritten -- there are no restrictions.

Preparation

The best way to prepare is to review your homework and labs. If there are any homework problems you did not understand or fully solve, please ask about them in office hours!

Question Types

Similar to the midterm:

- Short questions similar to the homework but easier.
- True/false, with a short statement to justify your answer.
- "Always, sometimes, never" questions.

Topics

Below are a list of things you should know, roughly catagorized into topics.

First half of course

We won't be specifically tested on material from the first half of the course.

Learning Theory

- What is a hypothesis class?
- What is the statistical learning model?

- What is PAC (probably approximately correctly learning)?
- What does it mean when we assume we are in the ``realizable setting''.
- What is the union bound? Be able to state and use it.
- You should be able to fully understand the proof of the main sample complexity theorem on Slide 19.
- Be able to bound the size of finite hypothesis classes like we did on the homework.

Kernel Methods

- What is a k-nn classifier?
- The inner product is a natural similarity measure between two data points, what are other similarity measures.
- What is a positive semidefinite kernel function? What are some examples?
- Be able to prove a simple kernel is positive semidefinite as we did on the homework.
- Know the reformulately versions of least squares regression and logistic regression that accomodate a non-linear kernel.

Support Vector Machines

- What is a seperating hyperplane? How is it mathematically defined? What is margin?
- Know the hard margin SVM objective. What is the definition of a support vector?
- Soft-margin SVM objective. How does changing the parameter C impact the final seperating hyperplane?
- How is the soft-margin objective related to logistic regresison?

Neural Networks Basics

- Understand why neural networks are so powerful (automatic feature learning).
- What are some examples of non-linearities? Why do neural networks always involve a non-linearity between layers? Because without a non-linearity, the layers could be collapsed into a single equivalent layer.
- Understand different diagrams of neural networks and what they mean.
- What is the multivariate chain rule? Be able to apply it.
- Be able to carry out a full-backprop calculation as in the homework.
- Understand the computational complexity of backprop (linear in number of neural network parameters).
- How does generalization in neural networks seem to differ from more traditional machine learning models? Neural networks will often overfit to the training data, but this won't necessarily hurt their test loss. Be able to recognize a convergence curve for a neural netowrk (e.g. Slide 48, Lecture 11) and differentiate it from that of a model line linear

regression.

Convolutional Neural Networks

- How is 1, 2, and 3D convolution defined? Be able to manually perform a simple convolution or to recognize the output of a convolution.
- How can convolutions be used for edge detection? For pattern matching?
- Understand that convolutional networks can be viewed as fully connected networks with additional constraints (some weights forced to equal zero, and weight sharing). Understand how this significantly reduced the number of parameters in the network.

Transfer Learning and Feature Learning

- What is the goal in transfer learning? What is "one-shot learning". How would you use a trained neural network for one task to help with learning for another related task?
- Definition of an autoencoder and autoencoder loss.
- What is the ``bottleneck'' in an autoencoder and why is it necessary.
- Waht are some possible applications of autoencoders? Feature extraction, data denoising, missing data imputation, data compression, generating realistic sythetic data.

Principal Component Analysis

- What is the linear autoencoder behind PCA?
- What is the connection between the autoencoder loss for PCA and low-rank approximation.
- What is the definition of the SVD. E.g. understnad what each of the three component matrices are, what it means for U and V to be orthogonal, etc.
- How do you find optimal weights for a linear autoencoder using the SVD.
- What are the `prinipcal components'' and what are the` loading vectors''.
- Be able to recognize when a dataset might naturally have a good low-rank approximation.
- Understand why the loading vectors approximately preserve distances and inner products of the original data.

Semantic Embeddings

- What is Latent Semantic Analysis? How does it obtain `word'' and` document'' vectors? Vectors for similar words should have high inner product (positive) and for dissimilar words should have low inner product (negative or near 0).
- What is the connection between LSA and factorizing the word co-occurance matrices? How can LSA be generalized to other similarity measures?
- The left/right singular vectors of the word co-occurance matrix are equal to the right

singular vectors of the term-document matrix.

- Why does the word co-occurance matrix used in LSA have a *symmetric* low-rank approximation? General similarity matrices will not.
- What are node embeddings? How is similarity defined via random walks in a graph?
- What is self-supervised learning and how does it differ from using an autoencoder to learn features?

Reinforcments Learning

- Understand the basic problem formulation + setup. What is the state, agents, reward function.
- What is a Markov Decision Process?
- How are the value function and Q function defined for an MDP an policy Π?