Format

- The purpose of the quals is to ensure that you have a breadth of knowledge across all the major topics in machine learning and its mathematical foundations.
- You should take your quals in your **second year** if you already have machine learning experience prior to starting your Ph.D. and your **third year** if you're new to machine learning.
- You should understand the basic idea behind each of the topics well enough to teach/explain the topic to a new CS Ph.D. student.
- The oral exam is 1.5 hours long. Three professors will each ask two questions, for a total of six questions.
- Each question will be based on the following types:
    - For some topic X:
        - Give a brief one minute tutorial answer to "What is X?"
        - Write the formal definition on the board, and perform some basic derivations.
        - Give concrete examples of its use.
        - Explain its significance.
    - Compare and contrast two or more topics (e.g., compare SVMs and neural networks).
- Some follow up questions might include
    - When would you use this approach in practice?
    - When would you use one approach over another?
    - What are the strengths/weaknesses of an approach?
    - Let's specialize to this particular example. What are the implications?
- Things that you're not responsible for
    - Detailed proofs.
    - Knowing the constants on bounds.
    - Solving pset-style problems.

Tips

- Start studying early.
- Practice answering questions and explaining topics on the whiteboard to a friend. Especially if you're unfamiliar with oral exam, your ability to communicate your knowledge and answer follow-up questions is at least as important as you knowing the material.

Topics

- Supervised learning
    - Logistic regression

- Support vector machines
- Neural networks
- Boosting (AdaBoost)
    - Rob Schapire's lecture notes on AdaBoost (the proof of Theorem, starting from the middle of page 4, is optional)
    - Alternatively, coordinate-descent interpretation
    - Some broader views (optional)
- Decision trees
- Random forests

- Kernels
    - Definition of kernel function
    - Types of kernels (linear, polynomial, RBF)
    - Reproducing Kernel Hilbert Space (RKHS)
- Deep learning
    - Convolutional neural networks
    - Autoregressive models
    - Recurrent neural networks
    - Attention, transformers
    - Generative adversarial networks
- Unsupervised learning
    - K-means
    - Expectation Maximization
    - PCA
    - Matrix factorization
    - Variational autoencoders
- Reinforcement learning
    - Markov decision process
    - Value iteration
    - Model-based RL
    - SARSA
    - Q learning
    - TD learning
    - Policy gradient
- Probability and statistics
    - Distributions
        - Gaussian
        - Poisson
        - Gamma
        - Dirichlet
        - Multinomial
    - Exponential family
    - Sufficient statistics
    - Maximum likelihood

- Asymptotic normality of M estimators
- Fisher information
- Information theory [Cover and Thomas, Elements of Information Theory, Chapter 2],
    - Entropy, conditional entropy
    - Mutual information
    - KL divergence
- Bias-variance tradeoff (for squared loss)
- Jensen's inequality
- Hypothesis testing
    - p-value, power
    - t-test
    - Bootstrap
    - Multiple hypothesis testing

- Optimization [see below, but appropriate references covering everything are Boyd and Vandenberghe's Convex Optimization and Duchi's Introductory Lectures on Convex Optimization, chapters 2--3]
    - Algorithms (including convergence rates for convex and strongly convex cases)
        - Gradient descent
        - Newton's method
    - Stochastic gradient method
    - Types of programs
        - Linear programming
        - Quadratic programming
        - Semidefinite programming
    - Lagrangian duality
        - Weak versus strong duality
        - Complementary slackness
        - KKT conditions
- Online learning [Hazan's Introduction to Online Convex Optimization, chapters 1--3, 6.1--6.4]
    - Regret setting
    - Online gradient descent
    - Exponentiated gradient
    - Multi-armed bandits
- Bayesian methods
    - Prior, likelihood, posterior
    - Basic posterior computations for conjugate priors (Gaussian, Dirichlet)
    - Marginal likelihood, model selection
    - Hypothesis testing
    - Bayesian mixture model
    - Latent Dirichlet Allocation
    - Approximate inference

        - Variational Bayes
        - Gibbs sampling
- Graphical models
    - Bayesian networks
        - Explaining away
        - d-separation
        - Parameter estimation
    - Markov networks
        - Partition function
        - Parameter estimation
    - Inference
        - Exact inference in trees via dynamic programming
        - Loopy belief propagation, Bethe free energy
        - Mean-field variational inference
        - Sampling
            - Monte Carlo integration
            - Importance sampling
            - Gibbs sampling
            - Metropolis-Hastings
            - Detailed balance (definition)
        - Sequential Monte Carlo / particle filtering
    - Structure learning
        - Chow-Liu trees
- Differential privacy [at the level of Chapters 1, 2, and 3.1--3.5.0 (i.e. pages 1-43) of Dwork and Roth's Algorithmic Foundations of Differential Privacy]
    - Basic mechanisms (e.g. randomized response, Laplace mechanism, sensitivity)
    - Composition of privacy (*not* advanced composition)
- Causality
    - Potential outcomes framework
    - Average treatment effect
    - Counterfactual inference / off policy batch RL (we can restrict to 1 time step if we want)
    - Propensity score matching / importance sampling
- Learning theory
    - PAC learning framework
    - Concentration inequalities
        - Hoeffding's inequality
    - Generalization bound for finite hypothesis class
    - Measures of complexity (definitions and basic examples)
        - VC dimension
        - Rademacher complexity OR metric entropy

Readings

These are general references which should cover most (but not all) the topics above. You are also encouraged to read more widely.

- CS229 lecture notes: basic machine learning
- CS229T lecture notes: online learning, kernel methods, uniform convergence, asymptotic statistics
- CS228 course notes: covers most of the graphical model concepts above. For a more in-depth treatment of loopy belief propagation and Bethe free energy, see chapters 3-4 of Wainwright and Jordan. For sequential Monte Carlo, see Doucet, de Freitas, & Gordon.
- Optimization
  - Stephen Boyd's convex optimization book
  - For optimization and subgradient algorithms, Chapters 2 and 3 of Duchi's "Introductory Lectures on Convex Optimization," or the subgradient method notes for ee364b (Boyd/Duchi)
  - Optional: Ryan Tibshirani's convex optimization course notes for optimization algorithm descriptions and convergence rates
- CS234 covers RL, bandits, and does batch RL which covers potential outcomes, average treatment effects, counterfactual inference and importance sampling.
- Sutton & Barto's reinforcement learning book

Classes

- CS221: artificial intelligence
- CS228: probabilistic graphical models
- CS229: machine learning
- CS229T: statistical learning theory
- EE364A, EE364B: convex optimization