Pre-Final Exam Review

CMPUT 296: Basics of Machine Learning

Textbook Ch.1 - Ch.12, &
Machine Learning Handbook Ch.7

Logistics

- **Assignment #3** due **Thursday December 3**
    - But there is no class on Dec 3
- **Final exam Friday December 18**
    - Covers ***all*** course material
    - Exam will be on eClass during a 24 hour period
    - Random spot checks scheduled starting the following week

Lecture Structure

1. Recap & Logistics
2. Exam structure and details
3. Learning objectives walkthrough
    - **Clarifying questions** are the point of this class
4. Other questions, clarifications

Final Exam Details

- The exam is **Friday, December 18** via **eClass**
- A practice final will be made available
- There will be a **6 hour** time limit for the exam
    - Starting at **any time** between 12:01am and 11:59pm Mountain time
    - It should *not* take anywhere near this long (I aimed for it to take **2 hours**)
- You may use a **single**, **handwritten cheat sheet** if you wish
- You may use a non-programmable calculator if you wish
- All course material is included

Exam Structure

- There will be **130 marks** total
- There will be **5-6** multi-part questions
    - **How** you got your answer will be the bulk of the marks
- There will be **no coding** questions
    - But you may be asked to **execute a few steps** of an algorithm
- Every question will be based on the **learning objectives** that we are about to walk through
- **There will be five marks for uploading a picture of your cheat sheet**

Bias-Variance Tradeoff

- Explain the implications of the **bias-variance decomposition** for estimators
- Explain what quantity is estimated by **linear regression**
- Describe the advantages and disadvantages of the MAP estimator for linear regression (Gaussian prior)
- Describe the bias-variance tradeoff for **reducible error**
- Explain how the choice of **hypothesis class** can affect the bias and variance of **predictions**

Logistic Regression

- Define linear classifier, sigmoid function, logistic regression
- Explain why **logistic regression** is more appropriate for binary classification than linear regression
- Describe how to **estimate** a logistic regression classifier's parameters
- Describe the advantages of the **MLE formulation** of logistic regression over a direct training cost minimization

Generalized Linear Models

- Define a **transfer function**
- Explain the purpose of a transfer function
- Explain how to choose an appropriate transfer function
- Define the **natural exponential family** of distributions
- Define a **generalized linear model**
- Explain the advantages of using a GLM

Monte Carlo Estimation

- Describe how to **estimate a random quantity** using a collection of samples
- Describe when estimating via sampling is appropriate
- Define **grid search**
- Describe how to **generate a random sample** using a CDF and a uniform random number generator
- Define **importance sampling**
- Explain when importance sampling is appropriate

Generalization Bounds

- Define **Rademacher complexity** and **empirical Rademacher complexity**
- Explain the implications of upper bounds on generalization error that include a Rademacher complexity term
  - You ***do not*** need to memorize the upper bounds expressions

Probability

- Define a **random variable**
- Define **joint** and **conditional probabilities** for continuous and discrete random variables
- Define **probability mass functions** and **probability density functions**
- Define **independence** and conditional independence
- Define **expectations** for continuous and discrete random variables
- Define **variance** for continuous and discrete random variables

Probability (2)

- Represent a problem probabilistically
- Compute joint and conditional probabilities
- Use a provided distribution
  - I will always remind you of the density expression for a given distribution
- Apply **Bayes' Rule** to derive probabilities

Estimators

- Define **estimator**
- Define **bias**
- Demonstrate that an estimator is/is not biased
- Derive an expression for the variance of an estimator
- Define **consistency**
- Demonstrate that an estimator is/is not consistent
- Justify when the use of a **biased estimator** is **preferable**

Estimators (2)

- Apply **concentration inequalities** to derive **error bounds**
- Apply the **weak law of large numbers** to derive error bounds
- Apply concentration inequalities to derive **confidence bounds**
- Define **sample complexity**
- Apply concentration inequalities to derive sample complexity bounds
- Explain when a given concentration inequality can/cannot be used

Optimization

- Represent a problem as an optimization problem
- Solve an analytic optimization problem by finding **stationary points**
- Define **first-order gradient descent**
- Define **second-order gradient descent**
- Define **step size** and **adaptive step size**
- Explain the role and importance of step sizes in first-order gradient descent
- Apply gradient descent to numerically find local optima

Parameter Estimation

- Describe the differences between **MAP**, **MLE**, and **Bayesian** parameter estimation
- Define the **posterior**, **prior**, **likelihood**, and **model evidence** distributions
- Represent a problem as parameter estimation
- Represent a problem as a formal prediction problem
- Define a **conjugate prior**

Prediction

- Represent a problem as a **supervised learning problem**
- Describe the differences between **regression** and **classification**
- Derive the **optimal classification predictor** for a given **cost**
- Derive the **optimal regression predictor** for a given cost
- Describe the difference between **discriminative** and **generative** models
- Describe the difference between **irreducible** and **reducible error**
- Describe the assumptions implied by a given error model

Linear Regression

- Represent a problem as **linear regression**
- Derive the **optimal predictor** for a linear model with squared cost and Gaussian errors
- Derive the computational cost of the **analytical** solution to linear regression
- Derive the computational cost of the **gradient descent** and **stochastic gradient descent** solutions to linear regression
- Represent a **polynomial regression** problem as linear regression
- Represent a **nonlinear regression** problem as linear regression

Generalization Error

- Describe the difference between **empirical error** and **generalization error**
- Explain why **training error** is a **biased estimator** of generalization error
- Define **overfitting**
- Describe how to **estimate generalization** error given a dataset
- Describe how to **detect overfitting**
- Apply k**-fold cross-validation** to select hyperparameters and/or features
- Apply **bootstrap resampling** to select hyperparameters and/or features

Generalization Error (2)

- Describe how to compare two models using **confidence intervals**
- Describe how to compare two models using a **hypothesis test**
- Describe how to compare two models using a **paired t-test**
- Define a p**-value**
- Define the **power** of a hypothesis test

Regularization

- Explain how to **avoid overfitting** using cross-validation
- Define a **hyperparameter**
- Define **regularization**
- Define the **L1 regularizer**
- Define the **L2 regularizer**
- Represent L2-regularized linear regression as **MAP inference**
- Explain how to use **regularization** to fit a model
- Describe the effects of the **regularization hyperparameter** λ

Other Questions?