Final Exam Review

CMPUT 267: Basics of Machine Learning

Chapters 1 - 11

Goal of these Slides

- Highlight key concepts to be tested
- Additionally highlight what I will not test
  - It is in the notes for your knowledge, but hard to directly test

Probability

- Understand the following concepts
  - **random variables**
  - **joint** and **conditional probabilities** for continuous and discrete random variables
  - **probability mass functions** and **probability density functions**
  - **independence** and conditional independence
  - **expectations** for continuous and discrete random variables
  - **variance** for continuous and discrete random variables

Probability (2)

- Know how to represent a problem probabilistically
- Use a provided distribution
  - I will always remind you of the density expression for a given distribution
- Apply **Bayes' Rule** to derive probabilities
- **Will not be directly tested:**
  - I will not expect you to know specific pdf and pmfs

Estimators

- Understand the following concepts
  - **estimators**
  - **bias**
  - **consistency**
  - how to show that an estimator is/is not biased
  - how to derive an expression for the variance of an estimator
  - how to show that an estimator is/is not consistent
  - when the use of a **biased estimator** is **preferable**

Estimators (2)

- Apply **concentration inequalities** to derive **confidence bounds**
- Define **sample complexity**
- Understand how concentration inequalities can be used to characterize the sample complexity of an estimator
- Explain when a given concentration inequality can/cannot be used
- **Will not be directly tested**
  - You do not need to know concentration inequality formulas

Estimators (3)

- Understand the **sample average estimator** and its properties
  - unbiased estimator, characterize variance
- Understand the **maximum likelihood estimator** (MLE)
- Understand the **MAP estimator**, and contrast to MLE
- **Will not be directly tested**
  - You will not need to derive parameters for MLE and MAP on the exam

Estimators (4)

- Understand that MAP and MLE are **point estimates**, and the **Bayesian** estimator maintains the full posterior p(w | D)
- Understand the role of **conjugate priors** priors
- **Will not be directly tested**
  - Do not need to know specific conjugate priors
  - Will not need to obtain credible intervals
  - Do not need to know the formula for posterior risk nor the Bayesian estimator that minimizes posterior risk

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
- **Will not be directly tested**
    - Specific stepsize adaptation algorithms

Prediction

- Describe the differences between **regression** and **classification**
- Understand the **optimal classification predictor** for a given **cost**
- Understand the **optimal regression predictor** for a given cost
- Describe the difference between **irreducible** and **reducible error**
- **Will not be directly tested**
  - Deriving optimal predictors
  - Multi-label vs multi-class classification

Linear Regression

- Derive the **optimal predictor** for a **linear model** with squared cost and Gaussian errors
- Derive the computational cost of the **gradient descent** and **stochastic gradient descent** solutions to linear regression
- Represent a **polynomial regression** problem as linear regression
- **Will not be directly tested**
  - Do not need to know the closed-form solution with matrices

Bias-Variance Tradeoff

- Explain the implications of the **bias-variance decomposition** for estimators
- 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**
- **Will not be directly tested**
  - Do not need to know the bias and variance formulas of the MLE and MAP estimators for linear regression

Logistic Regression

- Define linear classifier, sigmoid function, logistic regression
- Explain why **logistic regression** is more appropriate for binary classification than linear regression
- Understand that the objective (cross-entropy) and update underlying logistic regression is different from linear regression
- Understand that we estimate p(y | x), and predict $\arg\max_{y \in \{0,1\}} p(y \mid x)$
- **Will not be directly tested**
  - Knowing the specific logistic regression update
  - That the squared error results in a non-convex objective, unlike the cross-entropy

Generalization Error

- Describe the difference between **empirical error** and **generalization error**
- Explain why **training error** is a **biased estimator** of generalization error
- Describe how to **estimate generalization** error given a dataset
- Understand that we can use **statistical significance tests** to compare two models
- **Will not be directly tested**
  - Different ways to get samples of error
  - Specific statistical significance tests

Regularization

- Understand that regularization constrains the solutions to mitigate overfitting
- Understand **L1 regularization** and **L2 regularization**
- Understand that L2-regularized linear regression is the **MAP objective**
- Describe the effects of the **regularization hyperparameter** λ
- Understand that l1 regularization does feature selection
- **Will not be directly tested**
  - The Laplace distribution
  - Deriving the MAP solution