Final Exam Review

CMPUT 296: Basics of Machine Learning

Chapters 1 - 10

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 **confidence bounds**
- Define **sample complexity**
- Apply concentration inequalities to derive sample complexity bounds
- Explain when a given concentration inequality can/cannot be used

Estimators (3)

- Describe the **sample average estimator** and its properties
  - unbiased estimator, characterize variance
- Describe the **maximum likelihood estimator** (MLE)
  - show that the sample average is an MLE estimator, if estimating the mean of a normal distribution
- Describe the **MAP estimator**, and contrast to MLE

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**

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

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 **irreducible** and **reducible error**

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 **gradient descent** and **stochastic gradient descent** solutions to linear regression
- Represent a **polynomial regression** problem as linear regression

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

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
- Describe how to **detect overfitting**
- Apply **validation** to select hyperparameters

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

- 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