Midterm Review

CMPUT 296: Basics of Machine Learning

Textbook Ch.1 - 6

Midterm Details

- The content is from Chapters 1 - 6
  - Chapter 6 is Introduction to Prediction problems
  - Chapter 7 is Linear Regression. Exam does not cover linear regression
- The exam only covers what is in the notes

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
  - e.g., how likely was the outcome?
- 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 manipulate 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** **(poll)**

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

Optimization

- Represent a problem as an optimization problem
- Solve an analytic optimization problem by finding **stationary points** **(poll)**
- **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

- **Formalize a problem as a parameter estimation problem**
  - e.g., formalize modeling commute times as parameter estimation for a Poisson distribution, using maximum likelihood
- **Describe the differences between MAP, MLE, and Bayesian parameter estimation**
  - MAP $\min_{\theta} p(\theta \mid \mathscr{D})$ versus MLE $\min_{\theta} p(\mathscr{D} \mid \theta)$
  - Bayesian learns $p(\theta \mid \mathscr{D})$, reasons about plausible parameters
- Define a **conjugate prior**

Prediction

- 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
- Understand that the optimal predictor is different depending on the cost
- Describe the difference between **irreducible** and **reducible error**

$$\mathbb{E}[C] = \underbrace{\mathbb{E}\left[\left(f(X) - f^*(X)\right)^2\right]}_{\text{Reducible error}} + \underbrace{\mathbb{E}\left[\left(f^*(X) - Y\right)^2\right]}_{\text{Irreducible error}}$$

Reducible error

Irreducible error

Summary slide for Prediction

- **Supervised learning problem:** Learn a **predictor** $f : \mathcal{X} \to \mathcal{Y}$ from a dataset $\mathcal{D} = \left\{ (\mathbf{x}_i, y_i) \right\}_{i=1}^{n}$
  - $\mathcal{X}$ is the set of **observations**, and $\mathcal{Y}$ is the set of **targets**
- **Classification** problems have discrete targets
- **Regression** problems have continuous targets
- Predictor performance is measured by the **expected** $\text{cost}(\hat{y}, y)$ of predicting $\hat{y}$ when the true value is y
- An **optimal predictor** for a given distribution **minimizes** the expected cost
- Even an optimal predictor has some **irreducible error**.
  Suboptimal predictors have additional, **reducible error**

Any Questions?

- Switch to going over the practice midterm(s)