Pre-Quiz Review

CMPUT 296: Basics of Machine Learning

Textbook Ch.1 - §7.1

Logistics

1. **"In-class" quiz** Thursday Oct 8 (day after tomorrow!)
    - Covers all material through section 7.1
    - Quiz will be on eClass during a 24 hour period
    - Random spot checks scheduled starting the following week
2. **Thought questions #2** also due October 8
    - TQ#1 will be marked by the end of this week

Recap: Optimal Predictors

- **Supervised learning problem:** Learn a **predictor** $f : \mathscr{X} \to \mathscr{Y}$ from a dataset $\mathscr{D} = \left\{(\mathbf{x}_i, y_i)\right\}_{i=1}^{n}$
    - $\mathscr{X}$ is the set of **observations**, and $\mathscr{Y}$ 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**

Recap: Linear Models

A **linear predictor** has the form $f(\mathbf{x}) = w_0 + w_1 x_1 + \ldots + w_d x_d = \sum_{j=0}^{d} w_j x_j = \mathbf{w}^T \mathbf{x}$

Traditional approach: Find the linear predictor that minimizes squared error on the dataset (aka **Ordinary Least Squares**)

Probabilistic approach:

1. Assume **i.i.d. Gaussian noise**: $Y \sim \mathcal{N}(\omega^T \mathbf{x}, \sigma^2)$
2. Use MLE to estimate model from resulting **parametric family**
   $\mathscr{F} = \left\{ p(\cdot \mid \mathbf{x}) = \mathcal{N}(\mathbf{w}^T \mathbf{x}, \sigma^2) \mid \mathbf{w} \in \mathbb{R}^{d+1} \right\}$
3. Use the **optimal predictor** for the estimated model $\mathbf{w}^*$:
   $f^*(\mathbf{x}) = \mathbb{E}[Y \mid X = \mathbf{x}] = \mathbf{w}^T \mathbf{x}$

Lecture Structure

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

Quiz Details

- The quiz is **Thursday, October 8** via **eClass**
- There will be a **3 hour** time limit for the quiz
  - 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 **90 minutes**)
- You may use a **single**, **handwritten cheat sheet** if you wish
- You may use a non-programmable calculator if you wish
- Weeks 1 through 5 are included
  - Everything up to and including Linear Regression

Quiz Structure

- There will be **130 marks** total
- There will be **3-4** 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**

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