Final Exam Review

CMPUT 267: Basics of Machine Learning

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)

- 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
- Will not be directly tested: 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

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)
- 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 $p(y \mid x)$, and predict $\arg\max_{y \in \{0,1\}} p(y \mid x)$
- **Will not be directly tested**
  - That using the squared error results in a non-convex objective, unlike the cross-entropy

Bayesian linear regression

- The final exam will not cover this part specifically, but the idea of Bayesian estimation learned earlier in the course is potential material for the exam.

Other Questions?