

STA-395 INTRODUCTION TO MACHINE LEARNING – EXAM STUDY GUIDE

Section 1 – Fundamentals

- Supervised vs. Unsupervised Learning
 - o Be able to explain when/how methods like principal component analysis and clustering are used. You do not need to worry about the mathematical details of these methods.
- Classification vs. Regression
 - o Know different performance metrics for classification (prediction of a categorical label) and regression (prediction of a numeric outcome).
 - o Know how to set up and analyze a confusion matrix in terms of false positives, true positives, etc. You will be given the formula for any more sophisticated performance metric in terms of false positives, true positives, etc.
 - o Know the benefits, limitations, and use cases of metrics like classification accuracy, ROC AUC, F1 score, etc.
 - o Understand the conceptual differences between one-vs-many and one-vs-one, and macro-averaged vs. micro-averaged, handling of multi-label classification performance.
- Model Training, Evaluation, and Hyperparameter Selection
 - o Understand the importance of using separate training and testing sets, and the problems that arise from data leakage.
 - o Understand k-fold cross-validation, including how it works and how to express it using pseudo-code.
 - o Understand how grid search and cross-validation can be used to select hyperparameters and compare the performance of different models.
 - o Understand the bias-variance tradeoff and how it pertains to the hyperparameters of various models (those listed in the section below).

Section 2 – High-level understanding of algorithms

- Expect questions asking you to explain, describe, or show the predictions for the following methods (for both classification and regression tasks):
 - o K-nearest Neighbors
 - o Decision Trees
 - o Support Vector Machines
 - o Linear and Logistic Regression
 - o Random Forest
 - o Gradient Boosting

Section 3 – Mathematics of machine learning

- Expect one multi-part problem involving a structured model and optimization of parameters using a gradient descent algorithm.
 - o Know the general algorithm of gradient descent and how to express it using pseudo-code.
 - o Be familiar with the backpropagation algorithm and how it relates to gradient descent.