Module 3 – Machine Learning Fundamentals

Week 3 – February 3, 2025

BOSTON UNIVERSITY

BU **Faculty of Computing & Data Sciences**
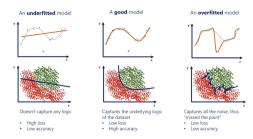
Week 2 Live Session: Agenda

- Where are we, where are we going?

- Bias-Variance? Why do we need to know that?

- General questions.

- Review of this week's lesson and resources.

- Let's code: Looking at HW 03.

Week 2: Where are we? Where are we going?

Week	Live Sessions Mon 7-8pm	HW Due Date Sun 11:59	Topic
Week 1	1/21 (Tues 8-9)	1/26	Introduction to Machine Learning
Week 2	1/27	2/2	Linear Regression
Week 3	2/3	2/9	Generalization, Errors, and Bias-Variance Tradeoff
Week 4	2/10	2/16	Training, Testing, and Validation Sets; Cross-Validation
Week 5	2/18 (Tues 8-9)	2/23	Feature Selection and Regularization
Week 6	2/24	3/2	Decision Trees
Week 7	3/3	3/9	Ensemble Methods: Bagging, Random Forests, Boosting Trees
Week 8	3/10	3/16	Classification
Spring Break			
Week 9	3/24		Variable Importance Measures
Week 10	3/31	4/6 (M 1)	Causal Inference
Week 11	4/7		Bias and Fairness
Week 12	4/14	4/20 (M 2)	Final Project Management
Week 13	4/22 (Tues 8-9)		Unsupervised Learning
Week 14	4/28	5/2 (Final)	Preview: Deep Learning

8 Weekly Homeworks

Project
  Milestone 1

Milestone 2

Final Report

You are here

BU Faculty of Computing & Data Sciences

Why is this week important?

o **Generalization is the "holy grail" of model building**—it ensures models perform well on unseen data, not just the training set. More than that, models that generalize effectively have learned the most important patterns in your data.

o **Understanding errors is key to model development**—distinguishing between **irreducible noise** and **model deficiencies** helps guide improvements and prevents misleading conclusions.

o **The bias-variance tradeoff provides a mathematical foundation for model complexity**—it helps separate **true patterns from random noise** and explains how models can fail in two ways:

- • **Underfitting:** Too simple, missing important relationships.

- • **Overfitting:** Too complex, learning noise instead of signal.

o **Model building is a balancing act**—most of your workflow will involve **tuning parameters** to find the right tradeoff between underfitting and overfitting. This week's material will give you the **tools and intuition** to manage these tradeoffs effectively and build models that generalize well.

An **underfitted** model — A **good** model — An **overfitted** model

Doesn't capture any logic / Captures the underlying logic of the dataset / Captures all the noise, thus "missed the point"
- High loss / Low loss / Low loss
- Low accuracy / High accuracy / Low accuracy

General Questions?

Before we look at the lesson and HW 02, do you have any general questions?

BU Faculty of Computing & Data Sciences

Overview of Lesson, Resources, and Homework

For the rest of the session, I'm going to screencast. Please raise your hand and ask questions at any point!

- Overview of Week 3 lesson and resources:

  - What is the purpose of instructor videos?

- Live Coding: Let's start Homework 03!

  - What is the point of each problem – what I am supposed to be learning?

  - How to solve the problems in the most Pythonic, machine learning-ish way?

  - What are the performance/interpretability/efficiency tradeoffs we should think about (if applicable)?

The End