

Introduction to Data Science
Homework 7: Due Friday October 19 at 2:00pm

Exercises:

1. Read Chapter 7 Mathematical Models in *The Data Science Design Manual* and respond to the following prompts:

- (a) What is modeling and what is a predictive model?
- (b) What is the take-home lesson from section 7.1.1?
- (c) Explain the bias-variance trade-off in your own words.
- (d) What is Bayes' theorem and how is it often applied? Why would you use Bayesian reasoning?
- (e) Compare and contrast each of the different pairs of modeling approaches:
 - Linear vs. Nonlinear models
 - Blackbox vs. Descriptive models
 - First-principle vs. Data-driven models
 - Stochastic vs. Deterministic models
 - Flat vs. Hierarchical models
- (f) What is a classification task in data science? Given an example.
- (g) What is a value prediction task in data science? Give an example.
- (h) What does it mean to “evaluate a model?”
- (i) Explain the expression:

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

- (j) Explain the expression:

$$\text{precision} = \frac{TP}{TP + FP}$$

- (k) Define *recall* and provide an expression for its value.
- (l) How do you compute an *F*-score? What is the role or use of an *F*-score in evaluating a model?
- (m) What is an ROC curve and how is it used?
- (n) Define absolute, relative and squared error. How is each computed?

- (o) Suppose that you have the following data: If we model the data by $y = 0.5x + 3$, then

x	1	2	3	4	5
y	3.4	4.1	4.2	4.7	5.3

what is the mean-squared error and the root mean-squared error for the model and data? Use R to make a plot of the data and the model predictions on the same set of axes. Write R functions that compute the mean-squared error and the root mean-squared error.