PROB 140

Probability for Data Science

Spring 2022

WEEK 14 STUDY GUIDE

The Big Picture

- For the bivariate normal, the conditional expectation is a linear function of the given variable, and hence is the same as the best linear predictor.
- The regression line can be written in multiple forms, one of which extends to the case of multiple regression.
- Prediction based on multiple predictors has familiar properties: There is a general formula for the best linear predictor, which is a natural extension of the formula for simple regression; and if the underlying distribution is multivariate normal then the best linear predictor is also the best among all predictors.
- The multiple regression model with normal errors is fundamentally important in data science. Properties of the estimated parameters lead to straightforward methods of inference.

Week At a Glance

Mon 4/25	Tue 4/26	Wed 4/27	Thu 4/28	Fri 4/29
	Lecture	Section	Lecture	Section
HW 13 Party 9am to noon **HW 13 Due** HW 14 (Due Mon 5/2)				
Focus on understanding HW 13	Work through Chapter 24	Skim Section 25.4	Work through Section 25.4	Work on HW 14

Reading, Practice, and Live Sessions

Book	Topic	Lectures: Prof. A.	Sections: GSIs	Optional Additional Practice
Ch 24, 25	**Towards Multiple Regression** - 24.4 writes the regression equation in multiple different ways, each one illuminating a different property and making it easier to understand the corresponding formulas in multiple regression - 25.1, 25.2, 25.3 extend the corresponding simple regression sections (24.1, 24.3, 24.4) to the multivariate case; we will just talk through these and not do the details - 25.4 introduces the multiple regression model most commonly used in data science	Tuesday 4/26 - MSE in simple regression; connection with the bivariate normal - The big picture of the multivariate case - The multiple linear regression model: understanding the assumptions	Wednesday 4/27 - Ch 24 Ex 3, 7, 6	None; focus on the homework
	Multiple Linear Regression - 25.4 continued: the estimates and their distribution under the model	Thursday 4/28 - Multiple linear regression model: parameter estimation and inference	Friday 4/29 - Ch 24 Ex 5 - Multiple regression model True/False	