Explanatory Notes for 6.390

Shaunticlair Ruiz (Current TA)

Fall 2022

The Linear Model

Now that we understand the problem of **regression**, and the concept of **optimizing** over it, we can pick a concrete example.

We want a function that can use information to **predict** outputs.

The Linear Model, 1-D

We'll start off small: we have one variable, and something we want to predict. And we'll pick the simplest pattern we can:

$$y = mx + b \tag{1}$$

A linear equation: m tells us how much our input affects our output. b accounts for everything unrelated to x: what is y when $x = 0$?

b and m are our parameters: that means they're part of Θ. We'll rename them θ_0 and θ_1.

$$h(x) = \theta_1 x + \theta_0 \tag{2}$$

The Linear Model, 2-D

We want to have **multiple** input variables: x will be a **vector**, not a number. So, for our above example, we'll **replace** x with x_1.

$$h(x) = \theta_1 x_1 + \theta_0 \tag{3}$$

The simplest way to include x_2 by just **adding** it. We have a scaling factor θ_1 for x_1, so we'll give x_2 its own **parameter**, θ_2:

If θ_1 is the "slope" for x_1, θ_2 is the "slope" for x_2.

$$h(x) = \theta_2 x_2 + \theta_1 x_1 + \theta_0 \tag{4}$$

The Linear Model, d-D

You can **expand** this to d dimensions by simply adding more terms:

This is the "dimension" of our input space: the **number** of input variables we have.

$$h(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + ... + \theta_d x_d \tag{5}$$

The Linear Model using Vectors

Here, we are **multiplying** components of x and θ together, then **adding**. This looks like a **dot product**:

$$h(x) = \theta_0 + \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_d \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix} \tag{6}$$

If we write this symbolically, we get:

$$h(x) = \theta_0 + \theta \cdot x \tag{7}$$

Unfortunately, we had to leave θ_0 out to make it work. θ is used for the parameters of our **dot product**, Θ is **all** parameters.

Notation 1

We represent the **parameters** of our **linear** equation as $\Theta = (\theta, \theta_0)$.

This formula looks similar to $y = mx + b$ again! Only this time, we have **vectors** instead.

We'll swap out the dot product for **matrix multiplication**: we'll use matrix multiplication a lot in this chapter, and course.

One benefit is that we can use **matrices** instead of just **vectors**!

Key Equation 2

The **linear regression** hypothesis is written as

$$h(x) = \theta^T x + \theta_0$$

Make sure you know what θ^T is: it's the **transpose**!

Remember that, when written out, this looks like:

$$h(x) = \begin{bmatrix} \theta_1 & \theta_2 & \theta_3 & \cdots & \theta_d \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_d \end{bmatrix} + \theta_0 \tag{8}$$

This is the **hypothesis class** of **linear hypotheses** we will reuse throughout the class.