Explanatory Notes for 6.390

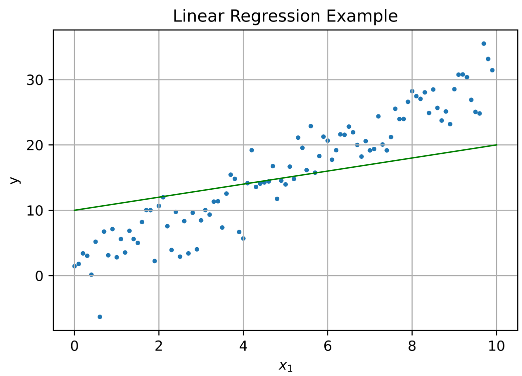
Shaunticlair Ruiz (Current TA)

Fall 2022

Regression Visualization

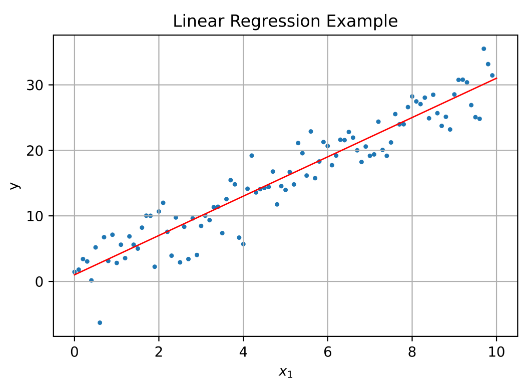
Visualizing our Model

With **one variable**, we've seen that our linear model simply turns into $\theta_1 x_1 + \theta_0$. As you'd expect, on a plot, this looks like a **line** in the **2D plane**.



This example of linear regression is not a great fit: $(\theta_0 = 10, \theta_1 = 1)$

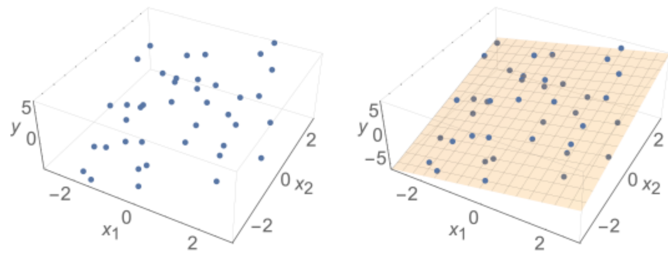
We're trying to get our line as **close as possible** to the points, hoping to find a linear pattern. We're **fitting** our line to the data.



This line is much better fitted to the data: $(\theta_0 = 1, \theta_1 = 3)$

What does this like if we have **two** variables? You need a 3D space, with 2 dimensions for the input.

Extending our line into a second dimension, we create a **plane**.



This plane is **fitted** the same way our line was. Notice that y is our **height**: this is the **output** of our regression.

Higher-dimension versions are hard to visualize. So, instead, we don't try, and call it a **hyperplane**.

Definition 1

A **hyperplane** is a **higher-dimensional version** of a **plane** - a **flat** surface that continues on forever.

We use it to represent our **linear** hypothesis for the purpose of **regression**.

The "**height**" ($(d+1)^{\text{th}}$ dimension) of this **plane** at a certain point represents the **output** of our **linear** hypothesis at that point.

Our line was a **1-D** object in a **2-D** plane. Our plane was a **2-D** object in a **3-D** space. So, our hyperplane is a d dimensional object in a $d+1$ dimensional space.

With this intuition, we can imagine our **hyperplane** as trying to get as **close** to all of the data points as it possibly can.

Another Interpretation

There's another, similar way to interpret our model

$$h(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + ... + \theta_d x_d \tag{1}$$

Before, we took θ_k as just an **extension** of the $mx + b$ formula: θ_k tells us how much x_k affects our output.

However, we can also think about the **relative** scale of each θ_k: if θ_2 is **larger** than θ_1, then x_2 has a **stronger** effect on the output than x_1.

We can say that x_2 **weighs** more heavily in our calculation: it has more say in the **result**.

Because of that, we sometimes call θ_k the **weight** for x_k.

Definition 2

A **weight** is a **parameter** that tells us how **strongly** a variable **influences** our **output**.

It is usually a **scalar** that we **multiply** by our variable.