Monday: Introduction to Supervised Learning

Agenda

This section should take you ~25 minutes

Specific Learning Outcomes

- I can explain the transition from descriptive to predictive analytics

Overview

Welcome to Module 2 of Core! This will be a major shift in topics, and we will leverage all our learnings from module 1 to transition to **predictive analytics.**

In Module 1, our focus was on **descriptive analytics.** What does the data tell us? Are there any outliers? What kind of distribution are we dealing with?

Building upon this knowledge, we will now start **predicting** new information. If we know demographic information about customers, as well as their purchasing habits, can we predict

how much a new person would purchase based on their demographic information?

If you have a few years of student data, covering mock exam results and national exam results, can we predict how well new students will do in the national exam if we know their mock exam results?

Throughout this module, we will focus on **supervised learning:** A series of algorithms that **learn** from established datasets in order to create **models.** You will need to gather, find, and clean up **training data set,** where you will have information on the variable you wish to predict (**the dependent variable**), as well as all the data you wish to use to make your prediction (**the independent variables**)

For our **purposes,** a **model** is like a function which can **predict the value of a dependent variable,** based on an input of one or more **independent variable.** Our typical workflow throughout this module will be like this:

- Identify and clean a data set
- Run descriptive analytics to identify patterns in the data, and decide which **supervised learning algorithm to use.**
- Create a **training data set** to create our model
- Assess how good our model is
- Perform our predictions

Main Resources

Introducing regressions:

The first family of algorithms we will study this week are called **regressions.** This might be familiar to some of you from high school/university: These algorithms try to identify the **line of best fit,** a line which minimizes how far away all the points in our graph are from it.

Fit at iteration 0
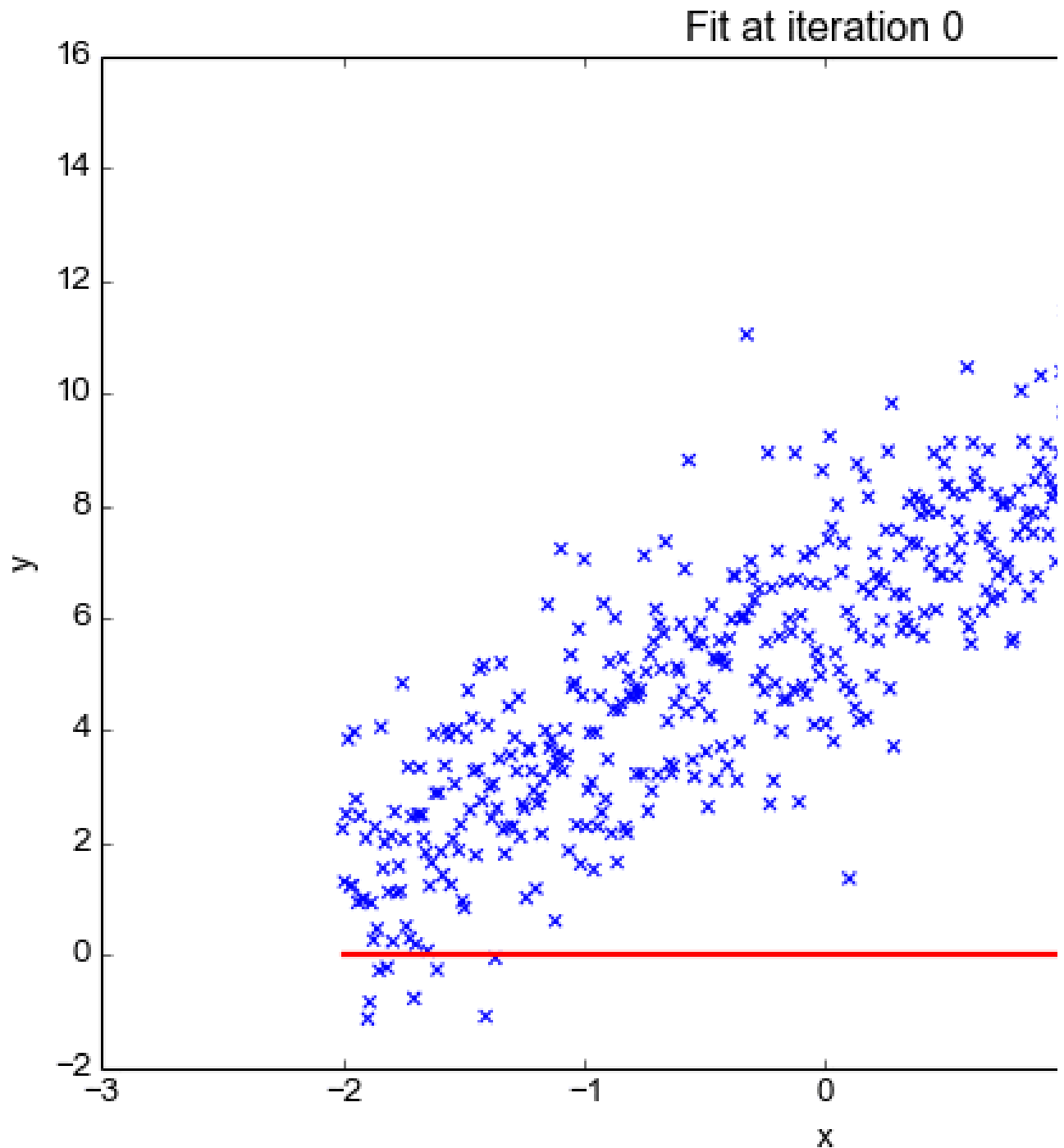
Suggested Follow-ups

- Introducing Supervised Learning. [**Link (https://www.youtube.com/watch?v=lg1nfPjrETc)**]

"A baby learns to crawl, walk and then run. We are in the crawling stage when it comes to applying machine learning." ~ Dave Waters