GERMS Machine Learning Reading Group
Meeting 1: Getting started

Paul Villanueva

Iowa State University

March 27, 2019

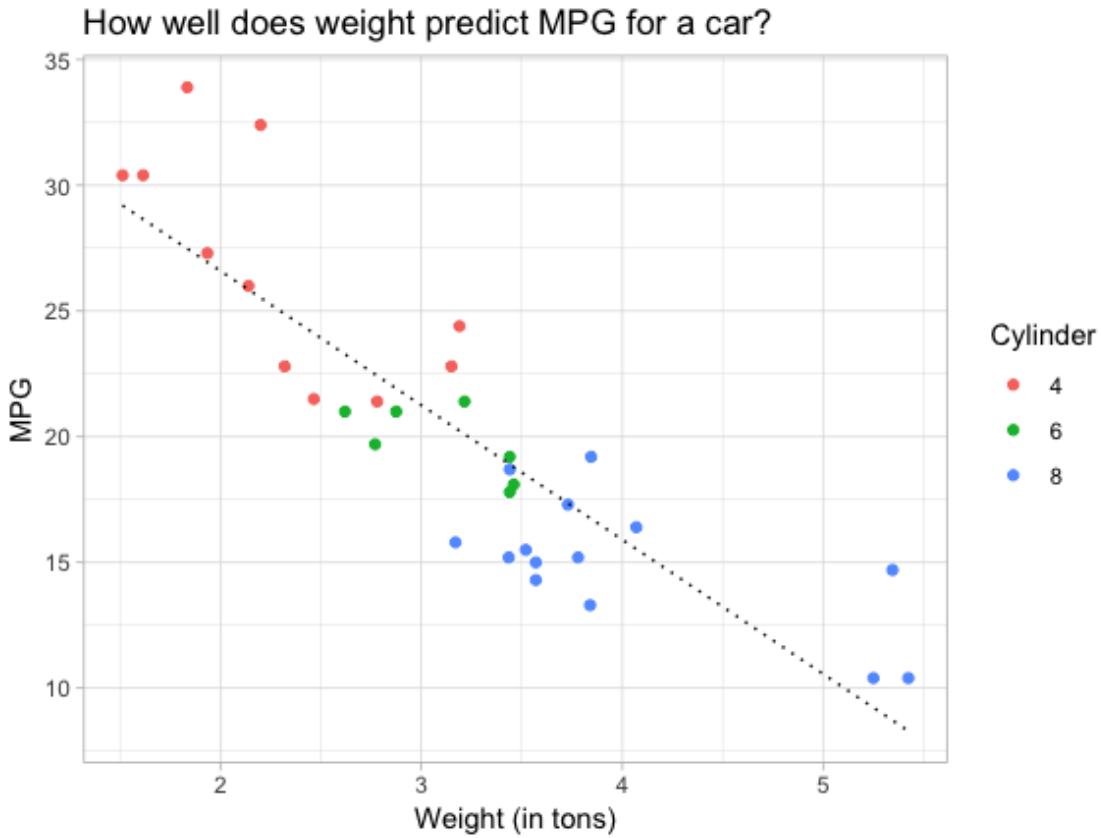
Outline

Introduction

- Machine learning is the process of teaching a computer to make predictions or decisions without being explicitly told how to do so.
- By feeding the algorithm data, it can pick up on patterns in the data to make inferences on new, unseen data.

Introduction

- A common example of this is linear regression:



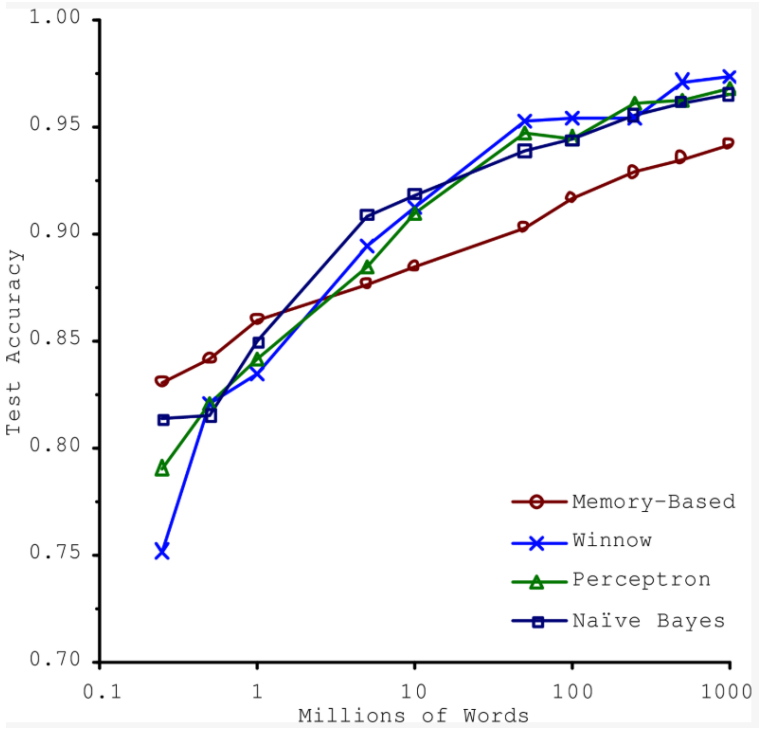
- The linear model is fit to the observed data, which we then use to predict future values.

Introduction

- Some other machine learning algorithms:
  - PCA, NMDS
  - Neural networks
  - k-Nearest Neighbors
  - Support vector machines
  - Random Forests
- "No Free Lunch" principle - there is no single best machine learning algorithm.
- Thus, all machine learning experiments involve testing various models in order to determine the best one.

Introduction

- All machine learning algorithms perform well as it sees more and more data.



- However, access to data is often limited, so knowing which models are appropriate is important

Introduction

- The general workflow for a machine learning project is:
  - Data curation (handling outliers, missing values, feature scaling, splitting datasets, etc.)
  - Exploratory data analysis (visualization, adding new features, etc.)
  - Model selection (choose a few models to work with and fit it to your training data)
  - Evaluate performance (how well did the model predict unseen values?
  - Tune hyperparameters (vary model parameters to improve performance)
  - Evaluate performance
  - Tune hyperparameters...
- We'll go over all these steps in weeks to come.

Next week

Two options for next time:

- Begin working through a machine learning project from beginning to end.
- Learn about the theory behind an algorithm (eg, neural nets)

Both options will have theoretical and applied aspects, but the first one will be more applied and the second one will be more theoretical.

What do you want to do?

A brief demonstration of `scikit-learn`

- There are many free packages in various languages to do machine learning. Popular ones are:
  - TensorFlow
  - Keras
  - Theano
  - Various packages within MATLAB
  - scikit-learn
- We'll begin with 'scikit-learn'. Go to the tutorial here.