Ensemble Models: Stacking

Hunter Glanz

OUTLINE

Foundational Machine Learning

- You've learned about:
  - Traditional Regression
  - Logistic Regression
  - K-Nearest Neighbors
  - Discriminant Analysis
  - Support Vector Machines
  - Tree-Based Methods

Foundational Machine Learning

- You've learned about:
  - Traditional Regression
  - Logistic Regression
  - K-Nearest Neighbors
  - Discriminant Analysis
  - Support Vector Machines
  - Tree-Based Methods

Remember *there's no free lunch*!

Ensemble Learning Strategies

- Ensemble learning refers to algorithms that combine the predictions from two or more models:
  - Let's team up!

Ensemble Learning Strategies

- Ensemble learning refers to algorithms that combine the predictions from two or more models:
  - Let's team up!
  - Near infinite number of ways to do this so we'll talk generally about three broad strategies:
    1. Bagging
    2. Stacking
    3. Boosting

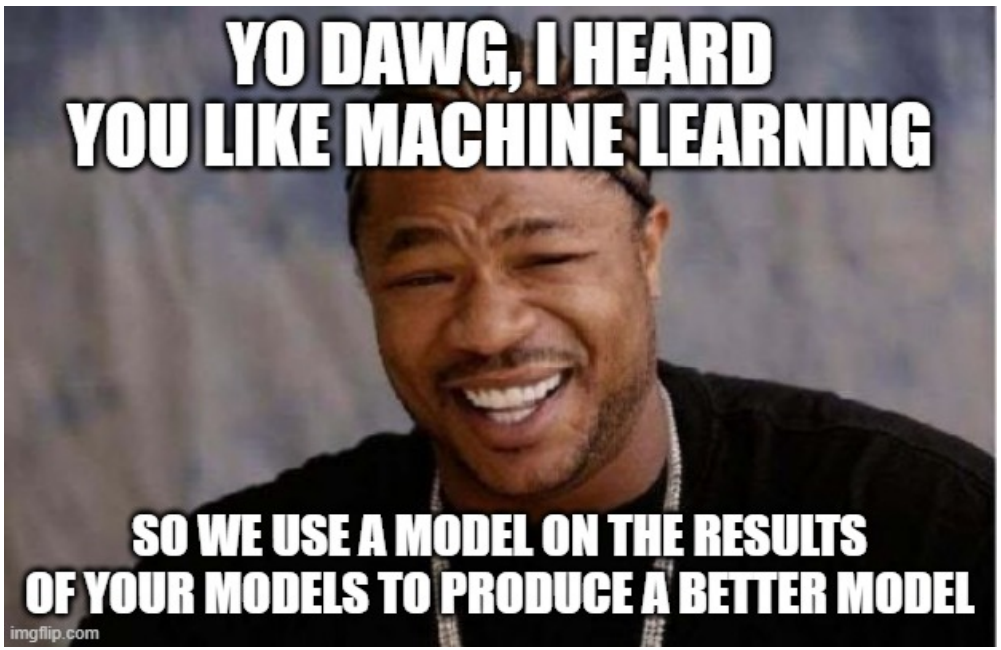
Ensemble Learning Strategies

- Ensemble learning refers to algorithms that combine the predictions from two or more models:
  - Let's team up!
  - Near infinite number of ways to do this so we'll talk generally about three broad strategies:
    1. Bagging
    2. Stacking
    3. Boosting

Today we will focus on **stacking**

Motivation

Given multiple machine learning models that are skillful on a problem, but in different ways, how do you choose which model to use (trust)?



The Stacking Approach

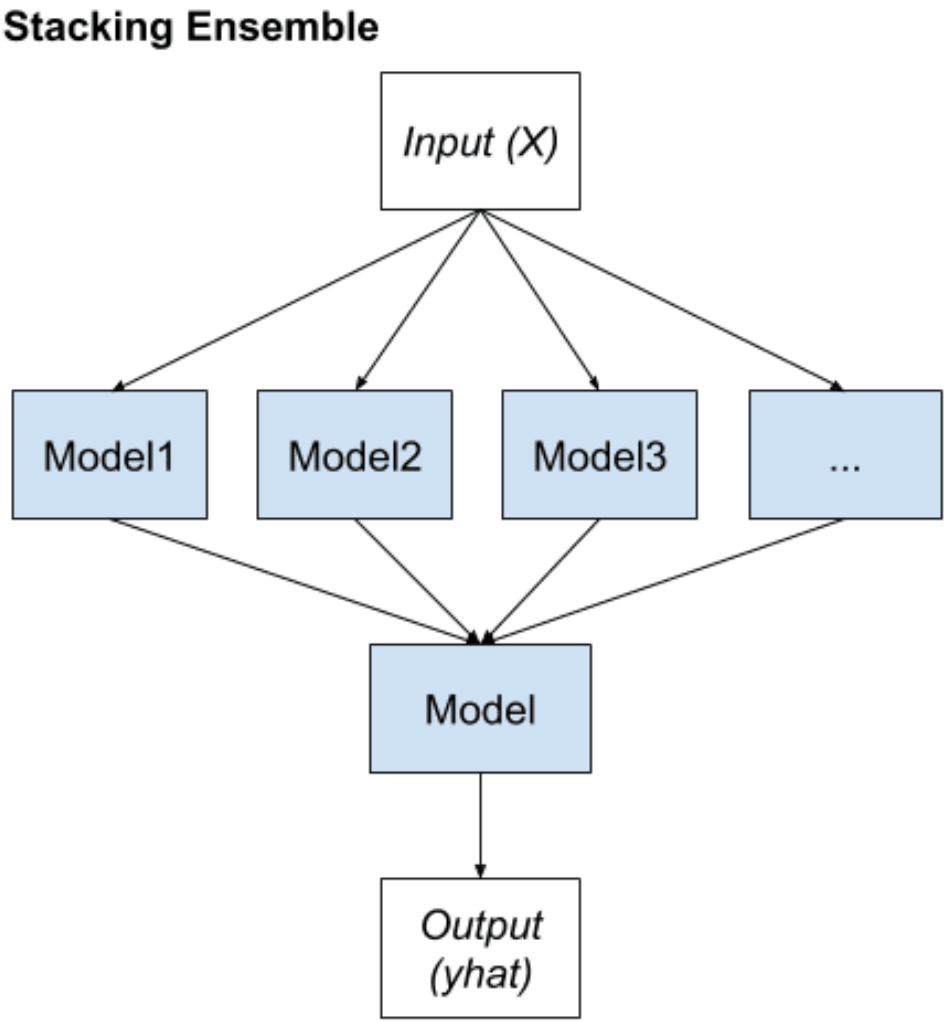
Use another machine learning model that learns when to use or trust each model in the ensemble

The Stacking Approach

Use another machine learning model that learns when to use or trust each model in the ensemble

- The base models are typically different (e.g. not all decision trees)
- The base models are fit on the same dataset (e.g. instead of samples of the training dataset)

The Stacking Visual



The Stacking Implementation

- **Level-0 Models (Base-Models)**: Models fit on the training data and whose predictions are compiled.
- **Level-1 Models (Meta-Model)**: Model that learns how to best combine the predictions of the base models.

The Stacking Implementation

- **Level-0 Models (Base-Models)**: Models fit on the training data and whose predictions are compiled.
- **Level-1 Models (Meta-Model)**: Model that learns how to best combine the predictions of the base models.

1. Base models are fit on training data. Predictions are made (and stored) on test data.
2. These test predictions are the inputs to the meta-model, where the outputs are the true response values in the test dataset.

One of the most common ways to set this up is via k-fold cross-validation

Stacking Notes

- Base-models are often complex and diverse (you can even use other ensemble methods as base-models).
- Meta-model is often simple, providing a smooth interpretation of the predictions made by the base models.
  - As such, linear models are often used as the meta-model, although this is not required.