Modeling the Cleansed Data

Mike West
MACHINE LEARNING ENGINEER

Module Overview

Core modeling approaches

Model categories

Model building steps

Model selection process

Demo: predicting Titanic survivability

Artificial Intelligence

AI is the top level container that holds all of machine learning. The science and engineering of making intelligent machines

Machine learning is a group of algorithms that give machines the ability to learn from experience. Machine learning is the parent to artificial neural networks

Artificial Neural Networks are algorithms that look for patterns in data that mimic the way the brain works

The Core Four

When the data is being used to predict a category, supervised learning is also called classification

When a value is being predicted, as with stock prices, supervised learning is called regression

Clustering is a machine learning technique that involves the grouping of data points

Association is a rule-based machine learning method for discovering interesting relations between variables

Core Model Building Steps

Fit

The training part of the modeling process

Predict

Classifying incoming data points

Evaluate

Determine the accuracy of the model

Model Selection Process

Starting

Data Amount
There is a minimum

Labeled Data

100K Samples
SGD vs SVC

Second Selection

Decision Tree
Results

Category Prediction

Labeled Data
Classification

Classifier Choice

SVC
Easily apply others

Demo

Import libraries

Load cleansed dataset

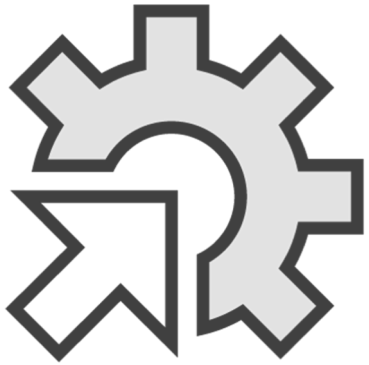
Build a Support Vector Classifier

Build a Decision Tree Classifier

Evaluate model results

Summary

- Traditional and deep learning
- Model categories
- Model building steps
- Model selection process
- The winning model



- The two core types
- Machine learning process
- Data wrangling importance
- Dataframes
- Data types
- NumPy for speed

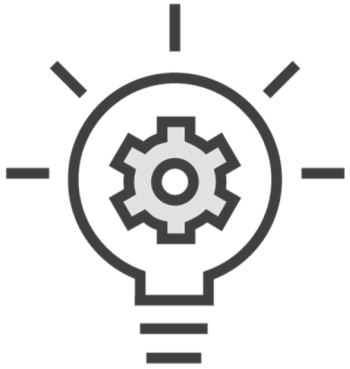
Series object

The array

Domain knowledge

Wrangling is process oriented

The AI ecosphere



The core four

SciKit-learn

Fit, predict and evaluate

Model selection

Winning model