Data Science Basics: Machine Learning

Information and practical exercises to add to your current toolkit or take the first step in launching a new career.

Welcome to Thinkful!

We teach tech skills that lead to fulfilling, high-paying careers.

Our students learn **in-demand** industry tools through **100% online programs** as they work toward a **job-ready portfolio** with the help of an **expert mentor**.

Let's get started.

Workshop Rundown

- ❏ Predictive Modeling Basics
- ❏ Assessing Models
- ❏ Improving Models
- ❏ Frequently Asked Questions
- ❏ Further Resources

Data Scientists

Data scientists take all the information collected by an organization and turn it into something valuable.

❏ Some specific steps in that process:
    ❏ Data Wrangling
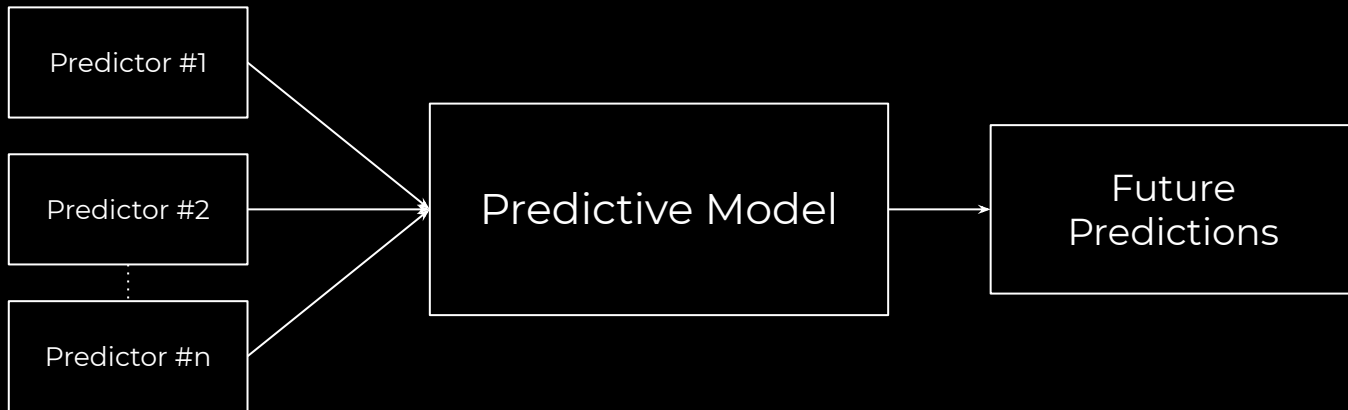   ❏ Analytics
    ❏ **Predictions**

Defining Machine Learning

" Machine learning algorithms provide computers with the ability to learn without being explicitly programmed — programming by example. "
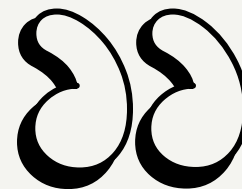
Practical Machine Learning

```
┌─────────────────┐
│   Predictor #1  │───────────┐
└─────────────────┘           │
                              ↓
┌─────────────────┐    ┌──────────────────┐      ┌──────────────┐
│   Predictor #2  │───→│ Predictive Model │─────→│    Future    │
└─────────────────┘    └──────────────────┘      │ Predictions  │
        ┆                                         └──────────────┘
┌─────────────────┐           ↑
│   Predictor #n  │───────────┘
└─────────────────┘
```

❏ Use past data to predict what will happen in the future.

❏ Predictors (variables) fed into model to train it to make predictions.

Pre-Modeling Tasks

❏ Data Cleaning
❏ Exploratory Data Analysis
❏ Data Selection and Transformation
❏ Data Set Splitting

Check out more Thinkful Webinars to get hands-on practice for these!
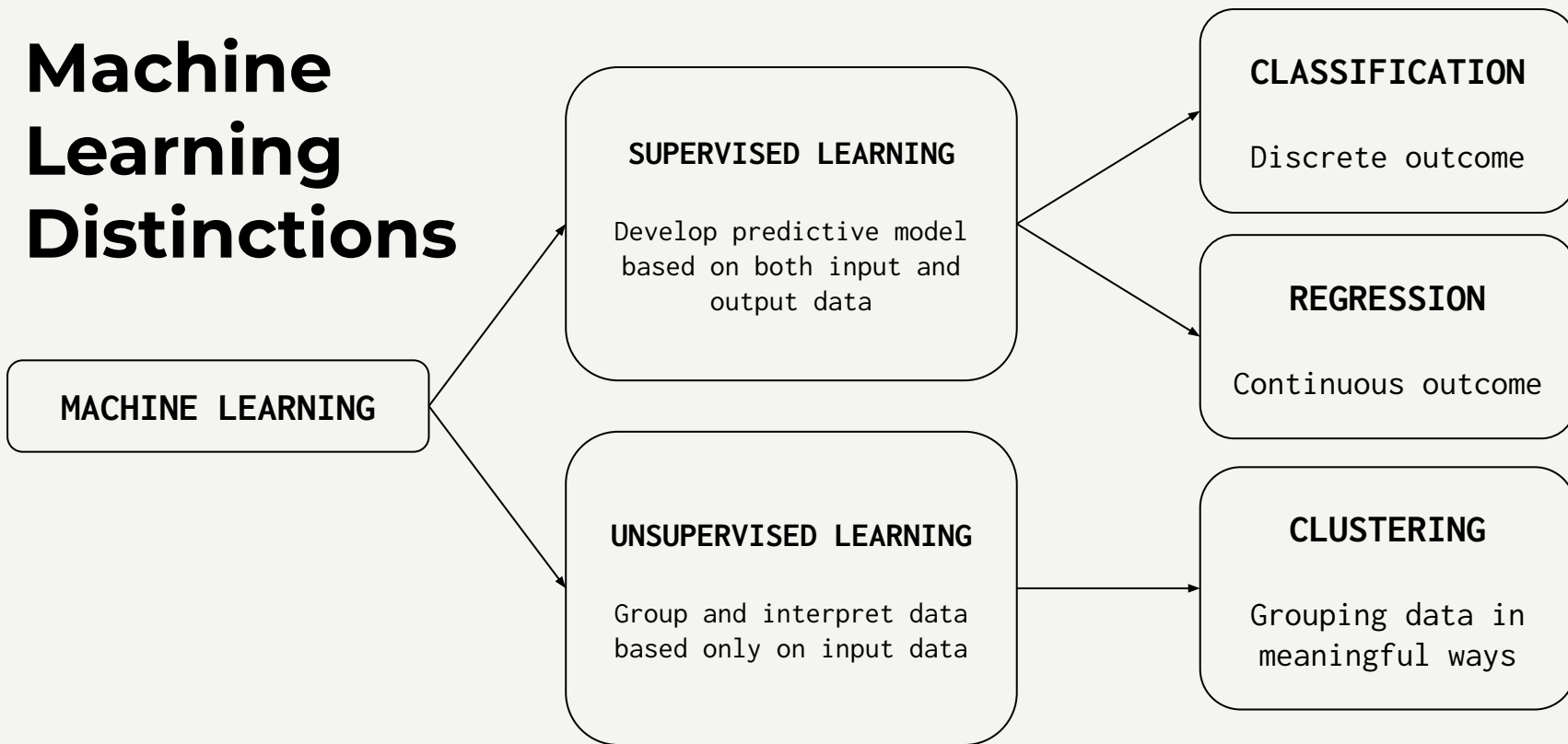
Key Modeling Tasks

How do we figure out the right model to use for our project?

❏ Know and clearly define the problem or question

❏ Choose models based on problem and data available

❏ Try a couple of different models to establish baselines

❏ Train the best performing model further

❏ Evaluate the model and report business implications

Machine Learning Distinctions

MACHINE LEARNING

SUPERVISED LEARNING

Develop predictive model based on both input and output data

UNSUPERVISED LEARNING

Group and interpret data based only on input data

CLASSIFICATION

Discrete outcome

REGRESSION

Continuous outcome

CLUSTERING

Grouping data in meaningful ways

Machine Learning Distinctions

CLASSIFICATION
(Supervised)
Discrete outcome

REGRESSION
(Supervised)
Continuous outcome

CLUSTERING
(Unsupervised)
Grouping data in meaningful ways

We have a bunch of data on whether or not a person defaulted on their loan. We want to predict whether or not a person should be extended a loan.

We have information about our volunteers, including information about their donation history, and we want to predict how much these people would donate.

We have data on our customers, such as the shows they watch on our streaming service. We want to create target groups of customers who will watch a new television show.

Evaluate Your Model

Classification Models

❏ Confusion Matrices
❏ Accuracy
❏ Precision

Regression Models

❏ r² Coefficient
❏ Root Mean Squared Error
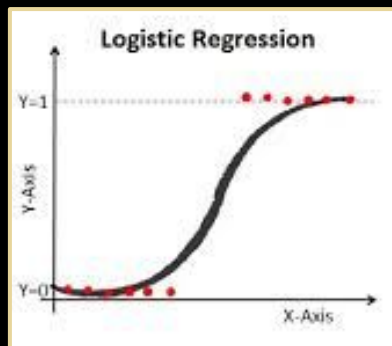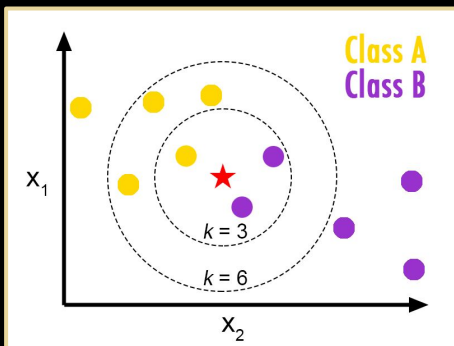
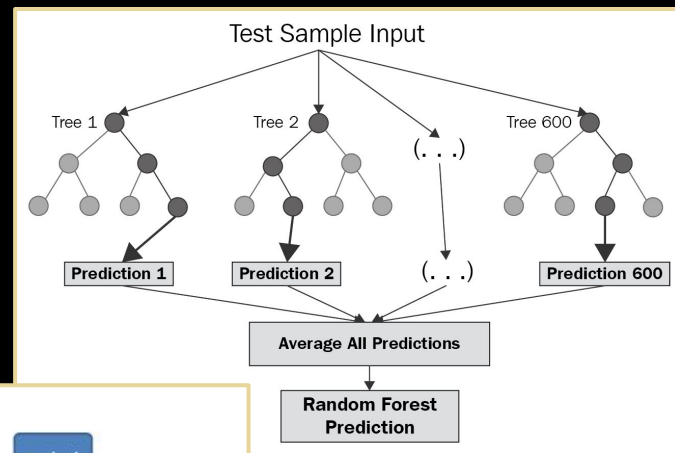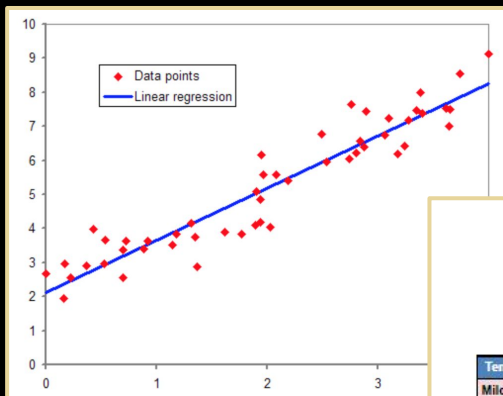Read this KDNuggets article for a more in-depth discussion.

Classification Examples

$$P(A|B) = \frac{P(B|A)\ P(A)}{P(B)}$$

using Bayesian probability terminology, the above equation can be written as
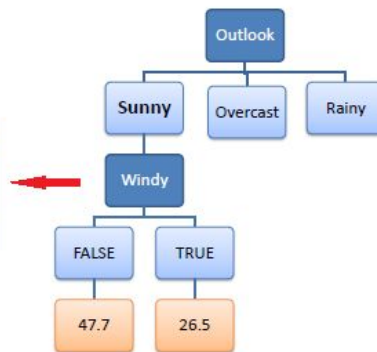
$$Posterior = \frac{prior \times likelihood}{evidence}$$
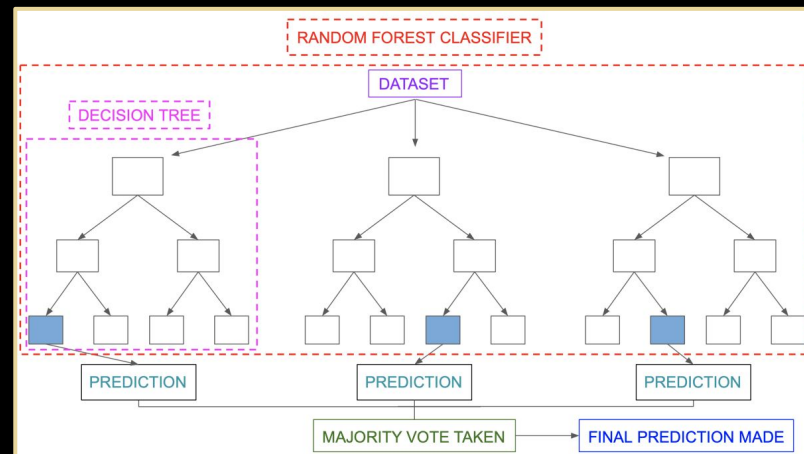
Class A
Class B

x_1

$k = 3$
$k = 6$

x_2

Logistic Regression

Y=1

Y-Axis

Y=0

X-Axis

RANDOM FOREST CLASSIFIER

DATASET

DECISION TREE

PREDICTION   PREDICTION   PREDICTION

MAJORITY VOTE TAKEN → FINAL PREDICTION MADE

Regression Examples

Predictive Model Example

Our Question

How much gross revenue will a new movie make?

Our Data

Information from past movies and their gross revenue, including:

❏ Actors and Directors
❏ Movie Length
❏ Budget

Our Goal

Use historical data to build a model predicting much revenue a movie will generate.

Machine Learning Distinctions

CLASSIFICATION
(Supervised)
Discrete outcome

REGRESSION
(Supervised)
Continuous outcome

CLUSTERING
(Unsupervised)
Grouping data in
meaningful ways

We have a bunch of data on past movies, including how much gross revenue a movie made. We want to predict how much gross revenue a movie will make.

The Notebook

Starter Code
bit.ly/ml-basics-colab

We'll be using a Google-hosted Python notebook called Colaboratory

❑ Click **File**
❑ Select **Save a Copy in Drive**
❑ This is your personal version of the notebook--let's get started!

Communicating Results

What do we do from here?

❏ What are important predictors to tell your boss that will help increase revenue for your movies?

❏ Explain to your boss *HOW* your chosen model is making predictions.

❏ **Stretch:** Create a <mark>visualization</mark> to show model performance

Common Questions

You might also be wondering

❏   What are the outcomes of your students
     for this field?
❏   How do I show my work to a potential
     employer?
❏   Is this course entirely online?
❏   What should I do from here?

THINKFUL

Take the First Step to A New Career

Anyone who's driven to change their future and achieve a high-earning career is able to enter the world's next workforce. We'll be by your side as you build the skills you need, with personal mentorship and an active, online community of students and educators.

Expand your career opportunities by breaking into tech. Chat with an admissions rep and we'll help you find the perfect fit.

Schedule a Call