Andrew Nystrom, Research Scientist @ Savvy Sherpa

What is Machine Learning?

A ridiculously high level overview

Why care?

- Do amazing things! (wait 3 slides)
- Very in demand
  - Data Science
  - Big Data
  - Data Mining
- Incredibly fun
  - A creative discipline
  - Combines math and CSci

What skills are needed?

- Linear algebra
- Calculus
- Probability & statistics
- Programming

Types of Learning

- Supervised learning
  - classification
  - regression
- Unsupervised learning
  - Clustering
  - Dimensionality reduction
- Reinforcement learning

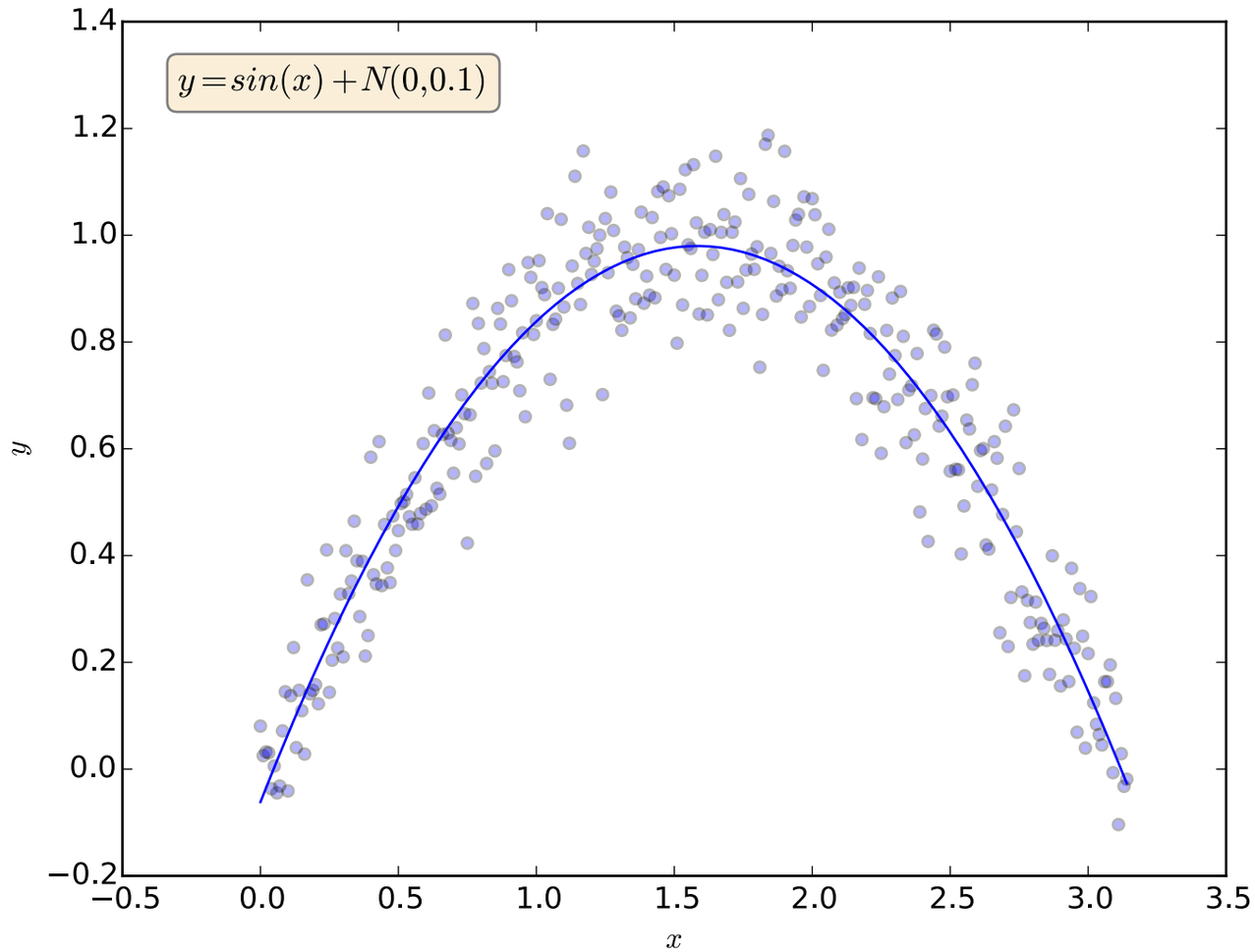
Supervised Learning Examples

- Spam filtering
- Market prediction
- Bone marrow donor/recipient matching
- Digit classification
- Image recognition
- Search engine query/document matching
- Product recommendation
- EEG reading to action prediction
- Voice to text (Siri)

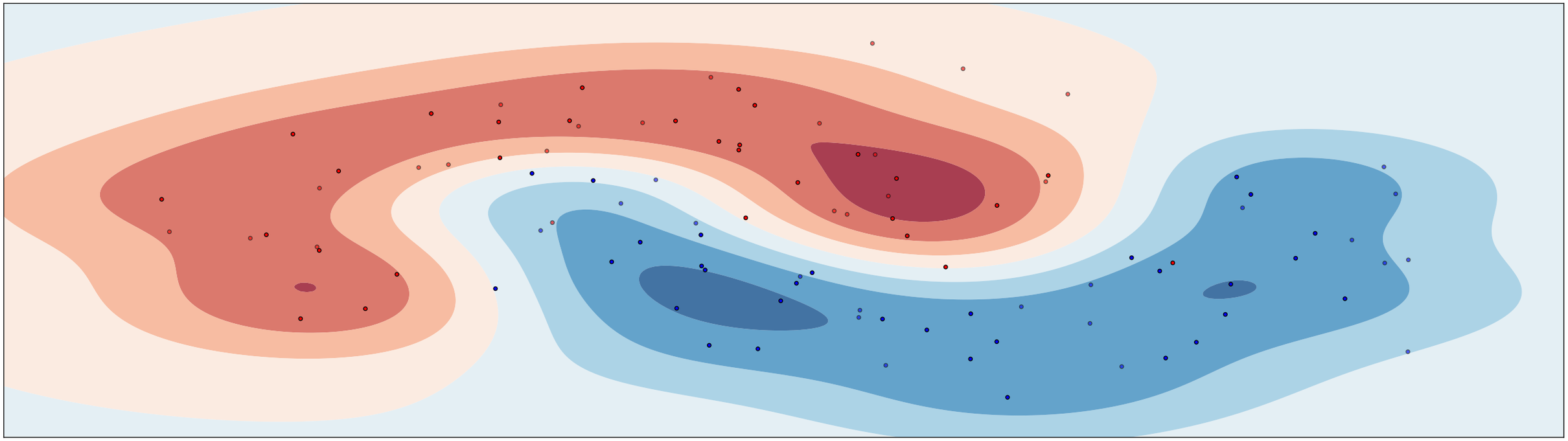
Definitions

- x: a vector that describes the data used to make a prediction.
- Could be…
  - An image
  - Previous values of a stock price
  - The text of a document
  - Components are called “features” - describe features of data
  - Get creative!
- y
  - class/label: what sort of “thing” x is (classification)
  - target: a real number being predicted (regression)

Regression Visualized

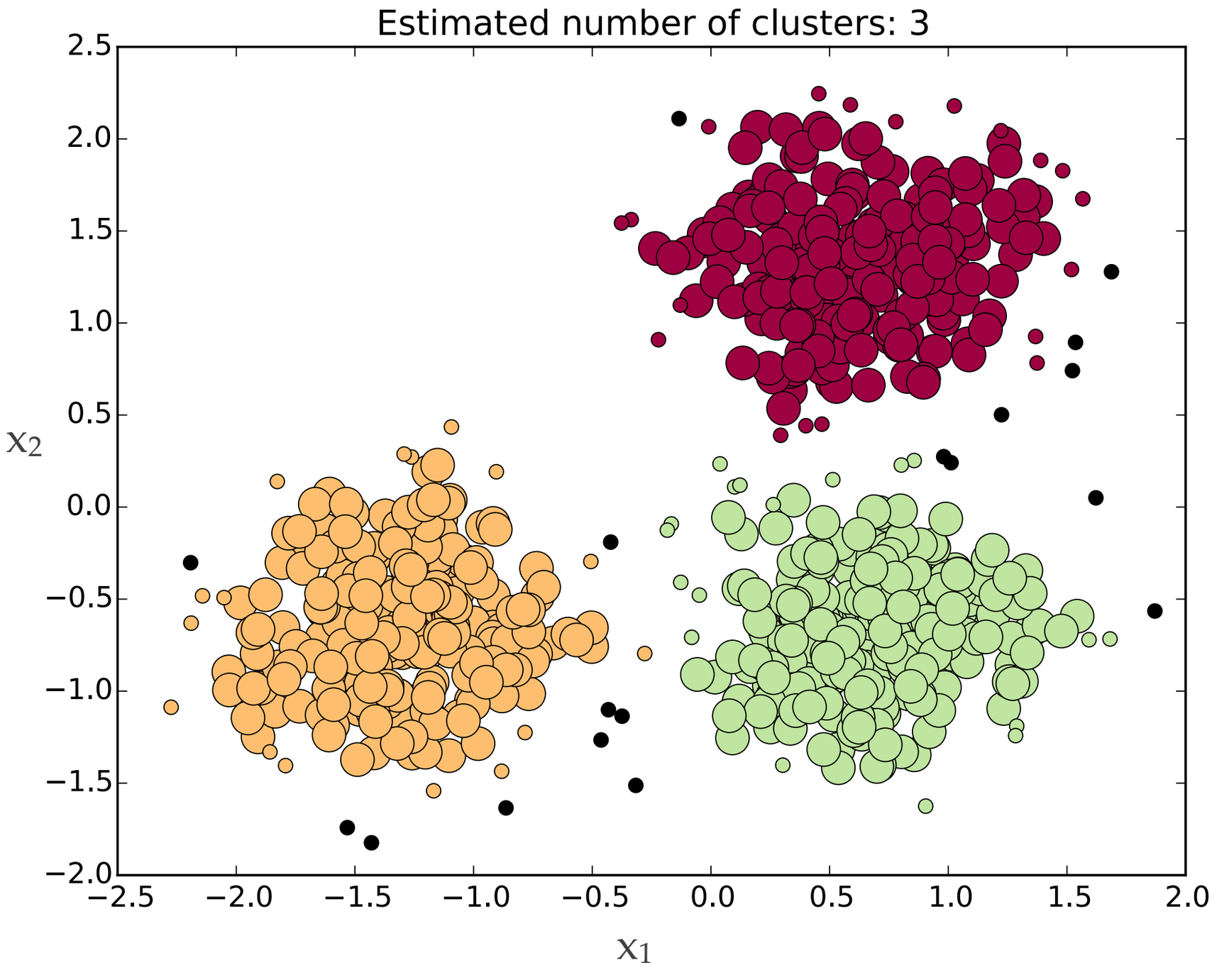


Classification Visualized



Color of point is true class label (y)
Color of region is classifier's certainty of class.

Clustering Visualized



Supervised Learning

- Get a bunch of examples of items and targets/labels
- Each item is a vector, x
- y is the target/label
- Learn a function f such that $f(x) \approx y$
- Now we can predict for x's we've not seen before!

Learning a function f

- Get some x, y pairs
- Define a function that links x to y
- Define the error and minimize it on the data
- Do it in a statistically sound way so f generalizes to unseen data

Example Model: Linear Regression

- Get some x, y pairs
- Define a function that links x to y
  - $y_{prediction} = wx+b$ *(learn w and b)*
- Define the error and minimize it on the data
  - error $= 1/n \cdot \sum(y_{actual} - y_{prediction})^2$ over all x, y pairs (mean squared error)
- This will find "good" values for w and b

Representation I: Stock Price Prediction

- Use linear regression to predict price of Google shares
- How to construct x?
  - Stock price from 5, 60, 1440 minutes ago
  - Time of day, day of week, day of month
  - Mentions of Google on Twitter
    - Positive, negative
- Each x consists of all of these, encoded as a vector.

Representation II: Spam Filtering

- doc1 = "Buy this thing for $9.99 plus tax and shipping" (spam)
- doc2 = "Someone died and you're the long lost heir…" (spam)
- doc3 = "Hey Andrew, want to grab a beer after work…" (ham)
- doc4 = "Andrew, I found a bug in your code…" (ham)
- …
- How to make x?
  - Vocabulary = all words in corpus (size $|V|$)
  - x is $|V|$ dimensional, each word in vocabulary has a component
  - For each word in doc, set component to 1. All others 0.
- This time use a classifier, not a regressor

Representation III: Image Clustering

- We have a ton of images. Group them by content
- E.g. nature photos go together, city photos, selfies, family photos, etc
- x could include:
  - number of colors
  - range of colors
  - number of faces
  - number of straight lines
  - number of distinct objects
- Apply clustering algorithm of choice

Sobel Edge Detection

Want to learn more?

- Free online courses
  - Andrew Ng's machine learning Coursera lectures
  - Patrick Winston's AI lectures
- University of Minnesota professors:
  - Vipin Kumar
  - Arindam Banerjee
- Software Libraries
  - Python: Scikit-learn
  - C++: mlpack, Shogun
  - Java: Weka
- Dataset repository: http://archive.ics.uci.edu/ml/
- Competitions: kaggle.com