Introduction to Machine Learning and Python

Shrey Gupta

Applied Machine Learning (HOUSECS 59-01), Duke University

August 29, 2018

Topics

- **Classification:** naive Bayes, support vector machines, kernel methods, and neural networks.
- **Regression:** spline interpolation and linear and polynomial regression.
- **Unsupervised learning:** mixture of Gaussians clustering.
- **Computer vision:** object detection via convolutional neural networks, feature extraction, edge detection, and processing methods.
- **Dimensionality reduction:** principal component analysis.
- **Evaluation of machine learning models.**

Prerequisites

- This is an *applied* course, but requires fundamental understanding of the algorithms and techniques being used.
- Prerequisites: basic fluency in programming and mathematics at the single-variable calculus level.

Grading

- **Attendance:** required to attend at least 11 (of 14) classes. 11 classes are lectures, and three are office hours.
- **Programming assignments:** required to complete all three assignments, with individual scores of 70% or greater.
- **Final paper:** 1500-word write-up detailing machine learning model from final programming assignment.
- Class is graded on a *satisfactory/unsatisfactory* basis.
- See syllabus for more detailed information.

What is Machine Learning?

- "Give computers the ability to learn without being explicitly programmed" (Arthur Samuel).
- Formal problem specification: "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E" (Tom Mitchell).

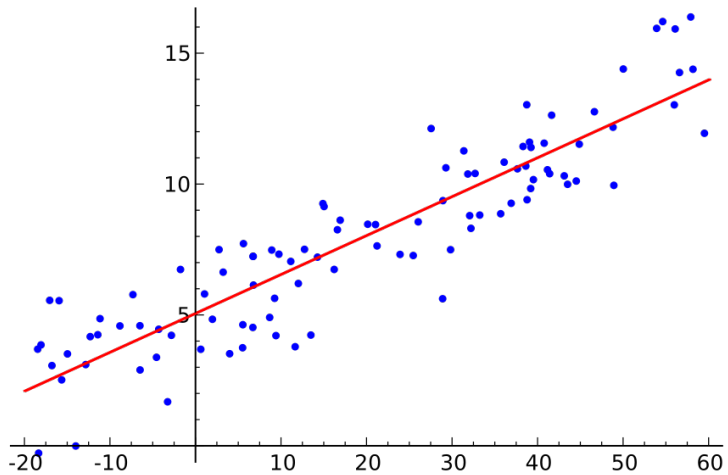
Terminology

- Grouped into two categories: *supervised* and *unsupervised* learning.
- Input data comes in *features* that describe each data point.
- *Training data*: data used to train algorithm (i.e. create model).
  - May include *noise* in the data.
- *Testing data*: untrained data that we seek insight on, potentially used to evaluate performance of model.
  - Ex: accuracy, mean squared error.

Supervised Learning

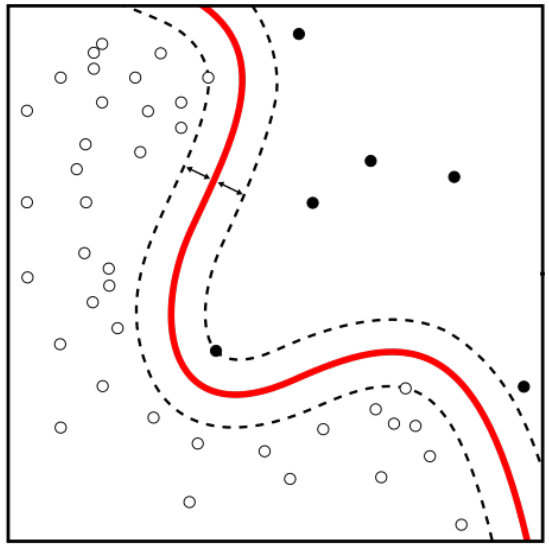
- Data (a subset from a larger distribution) is labeled, and we attempt to generalize to (predict) the larger distribution.
- Regression: predicts a continuous value output.
- Classification: predicts a discrete class output.

Regression: Examples



- Given data about square footage, age, zip code, and housing demand, predict the selling price of a house.
- Predict the percentage increase or decrease in the price of an equity.

Classification: Examples



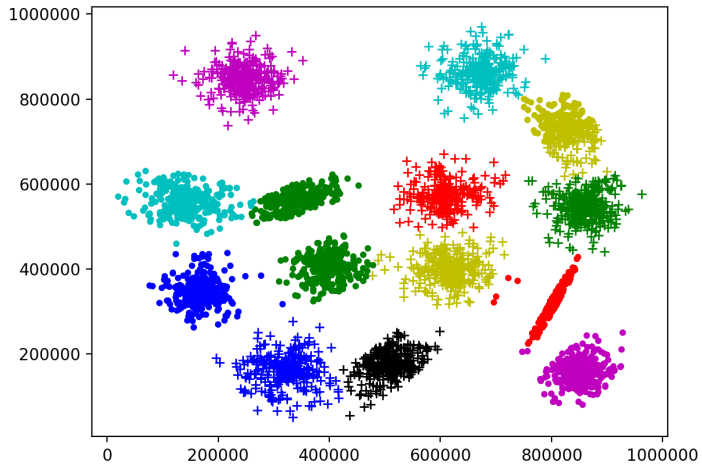
- Given data about temperature, humidity, and wind speed, predict whether it will be sunny, cloudy, or raining.
- Predict whether the price of an equity will increase or decrease.

Image source: Wikipedia

Unsupervised Learning

- Data is unlabeled (no "ground truth").
- Problems: clustering, density estimation, and pattern detection.

Clustering: Examples



- Given consumption data, partition the consumers into market segments.
- Given several news articles (and their text), group them based on similarity.

Notation

$$X = \begin{bmatrix} x_1^{(1)} & x_1^{(2)} & x_1^{(3)} & \dots & x_1^{(m)} \\ x_2^{(1)} & x_2^{(2)} & x_2^{(3)} & \dots & x_2^{(m)} \\ \vdots & & & \ddots & \vdots \\ x_n^{(1)} & x_n^{(2)} & x_n^{(3)} & \dots & x_n^{(m)} \end{bmatrix}, y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

- Data is stored in matrices and vectors.
- Given n (training) data points and m features (per data point).
- For supervised learning, given labeled data vector y.

Algorithm: Supervised Learning

$$X_{test} = \begin{bmatrix} x_1^{(1)} & x_1^{(2)} & x_1^{(3)} & \dots & x_1^{(m)} \\ x_2^{(1)} & x_2^{(2)} & x_2^{(3)} & \dots & x_2^{(m)} \\ \vdots & & & \ddots & \vdots \\ x_k^{(1)} & x_k^{(2)} & x_k^{(3)} & \dots & x_k^{(m)} \end{bmatrix}, \hat{y}_{test} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_k \end{bmatrix}$$

- Given k testing data points and m features (per data point).
- $\hat{y}_{test} = f(X_{test})$ contains *predictions* of the supervised learning algorithm, where $f(\cdot)$ is learned by the algorithm.

Algorithm: Unsupervised Learning

$$X_1 = \begin{bmatrix} x_i^{(1)} & x_i^{(2)} & \dots & x_i^{(m)} \\ x_j^{(1)} & x_j^{(2)} & \dots & x_j^{(m)} \\ \vdots & & \ddots & \vdots \\ x_k^{(1)} & x_k^{(2)} & \dots & x_k^{(m)} \end{bmatrix}, \dots, X_p = \begin{bmatrix} x_q^{(1)} & x_q^{(2)} & \dots & x_q^{(m)} \\ x_r^{(1)} & x_r^{(2)} & \dots & x_r^{(m)} \\ \vdots & & \ddots & \vdots \\ x_s^{(1)} & x_s^{(2)} & \dots & x_s^{(m)} \end{bmatrix}$$

- Partitions data into p clusters (based on some similarity measure).
- Algorithm may have some method to classify new data points into clusters.

Python

- We'll be using *Python 3.x* throughout the course.
- Libraries and frameworks: NumPy, SciPy, Pandas, Matplotlib, SciKit, TensorFlow, and NLTK.
- Today's (Jupyter) notebook will ensure all of the packages are downloaded and work through an introduction of Python.