Big Data Technology

Paul Rad, Ph.D.

Associate Professor
Information Systems and Cyber Security, College of Business School
Electrical and Computer Engineering, College of Engineering

Outline

- **Logistics**
- **Summary and Review Previous Sections**
- **Machine Learning**

A Few Quotes

"A breakthrough in machine learning would be worth ten Microsofts" (Bill Gates, Chairman, Microsoft)

"Machine learning is the next Internet"
(Tony Tether, Director, DARPA)

Machine learning is the hot new thing"
(John Hennessy, President, Stanford)

"Machine learning is going to result in a real revolution" (Greg Papadopoulos, CTO, Sun)

A Short History of Machine Learning

1950 — **Alan Turing** creates the “Turing Test” to determine if a computer has real intelligence. To pass the test, a computer must be able to fool a human into believing it is also human.

1952 — **Arthur Samuel** wrote the first computer learning program. The program was the game of checkers, and the IBM computer improved at the same the more it played studying which moves made up winning and incorporating those moves into its program.

1957 — **Frank Rosenblatt** designed the first neural network for computers (the perceptron), which simulate the thought processes of the human brain.

1967 — The “nearest neighbor” algorithm was written, allowing computers to begin using very basic pattern recognition. This could be used to map a route for traveling salesmen, starting at a random city but ensuring they visit all cities during a short tour.

1979 — Students at Stanford University invent the “Stanford Cart” which can navigate obstacles in a room on its own.

1981 — **Gerald Dejong** introduces the concept of Explanation Based Learning (EBL), in which a computer analyses training data and creates a general rule it can follow by discarding unimportant data.

1985 — **Terry Sejnowski** invents NetTalk, which learns to pronounce words the same way a baby does.

A Short History of Machine Learning

1990s — Work on machine learning shifts from a knowledge-driven approach to a data-driven approach. Scientists begin creating programs for computers to analyze large amounts of data and draw conclusions — or "learn" — from the results.

1997 — IBM's Deep Blue beats the world champion at chess.

2006 — Geoffrey Hinton coins the term "deep learning" to explain new algorithms that let computers "see" and distinguish objects and text in images and videos.

2010 — The Microsoft Kinect can track 20 human feature at a rate of 30 times per second, allowing people to interact with the computer via movements and gestures.

2011 — IBM's Watson beats its human competitors at Jeopardy.

2011 — Google Brain is developed, and its deep neural network can learn to discover and categorize objects much the way a cat does.

2012 – Google's X Lab develops a machine learning algorithm that is able to autonomously browse YouTube videos to identify the videos that contain cats.

A Short History of Machine Learning

2014 – Facebook develops DeepFace, a software algorithm that is able to recognize or verify individuals on photos to the same level as humans can.

2015 – Amazon launches its own machine learning platform.

2015 – Microsoft creates the Distributed Machine Learning Toolkit, which enables the efficient distribution of machine learning problems across multiple computers.

2015 – Over 3,000 AI and Robotics researchers, endorsed by Stephen Hawking, Elon Musk and Steve Wozniak (among many others), sign on open letter warning of the danger of autonomous weapons which select and engage targets without human intervention.

2016 – Google's artificial intelligence algorithm beats a professional player at the Chinese board game Go, which is considered the world's most complex board game and is many times harder than chess. The AlphaGo algorithm developed by Google DeepMind managed to win five games out of five in the Go competition.

Machine Learning Definition

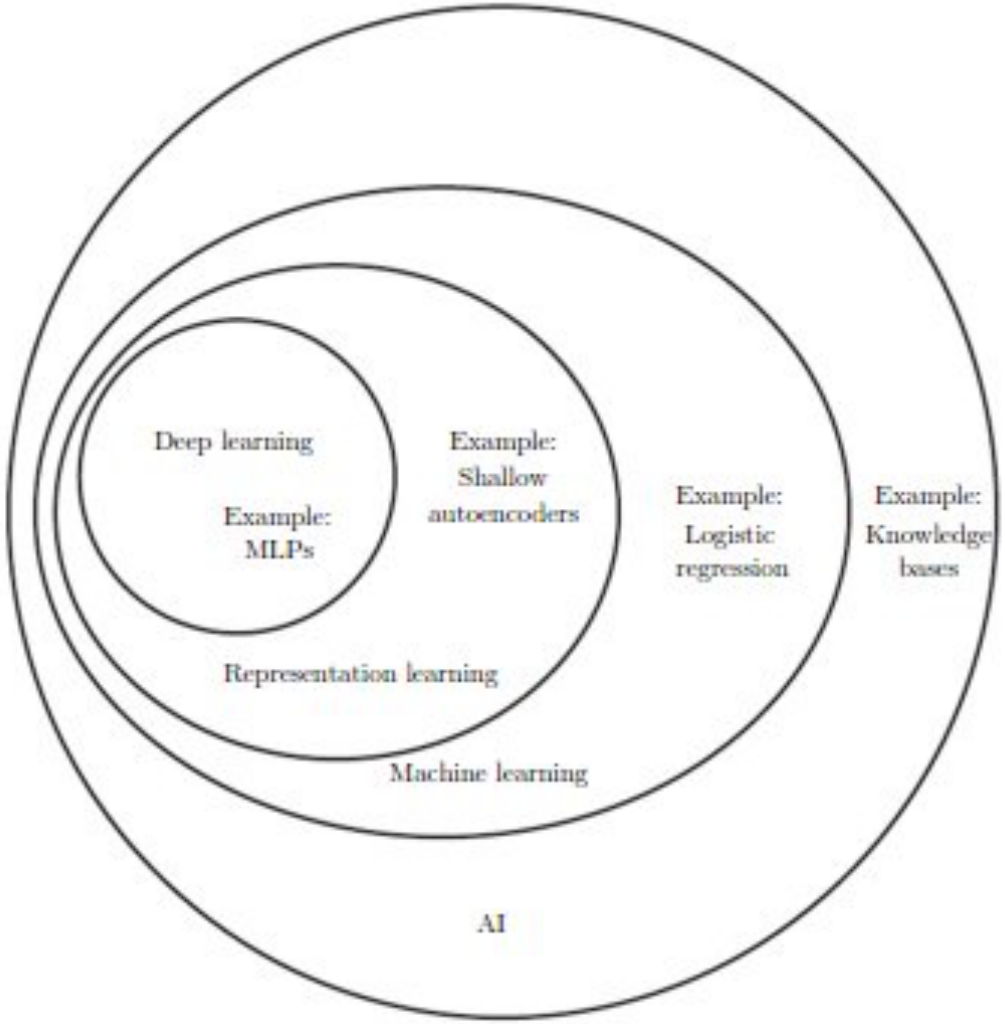
A machine learning algorithm is an algorithm that is able to learn from data.

Tom Mitchell (1997) provides the definition:

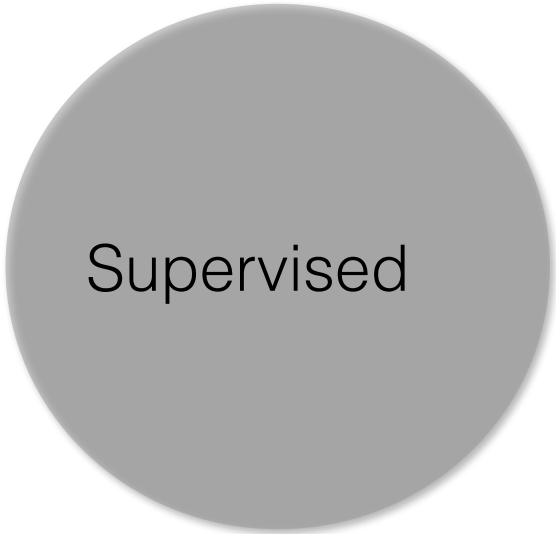
"A computer program is said to learn from

experience *E* with respect to some class of task *T* and performance measure *P,*

if its performance at tasks in *T*, as measured by *P*, improves with experience *E*."



3 Types of Learning



- Learning from labeled data
- E.g., Spam classification

- Classification
- Regression
- Ranking

- Discover structure in unlabeled data
- E.g., Document clustering

- Clustering
- Hidden Markov Models

- Learning by “doing” with delayed reward
- E.g., Chess computer

Supervised Learning

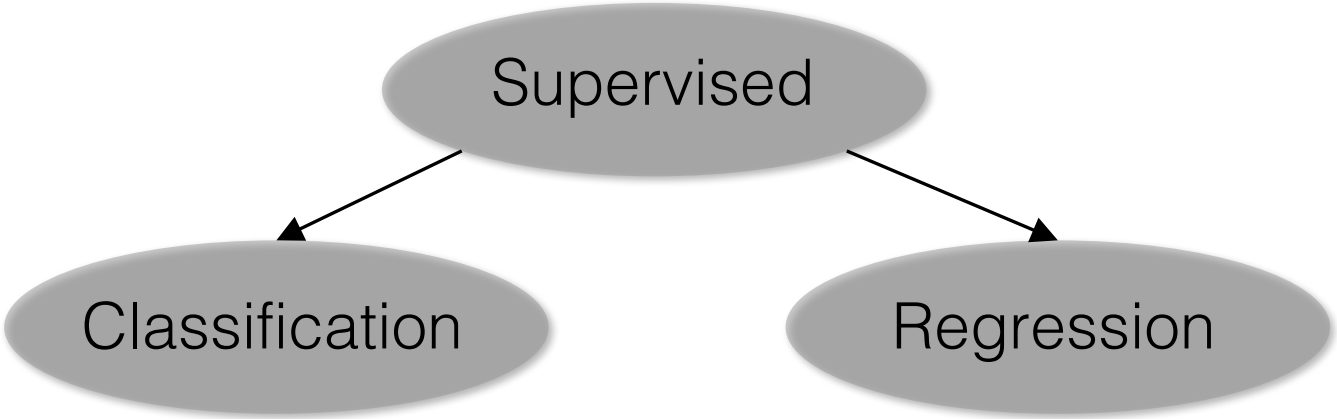
Given examples of a function *(X, F(X))*

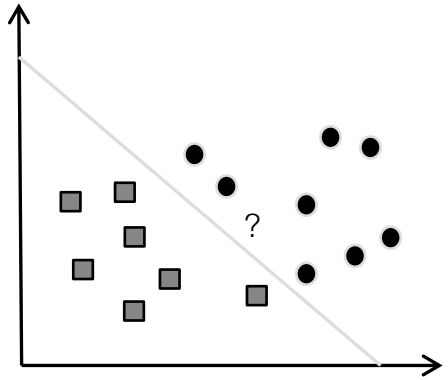
Predict function *F(X)* for new examples *X*

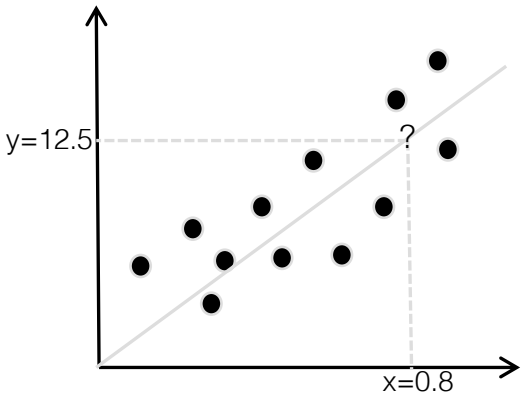
Discrete *F(X)*: Classification

Continuous *F(X)*: Regression

F(X) = Probability(*X*): Probability estimation







Regression and Classification Examples

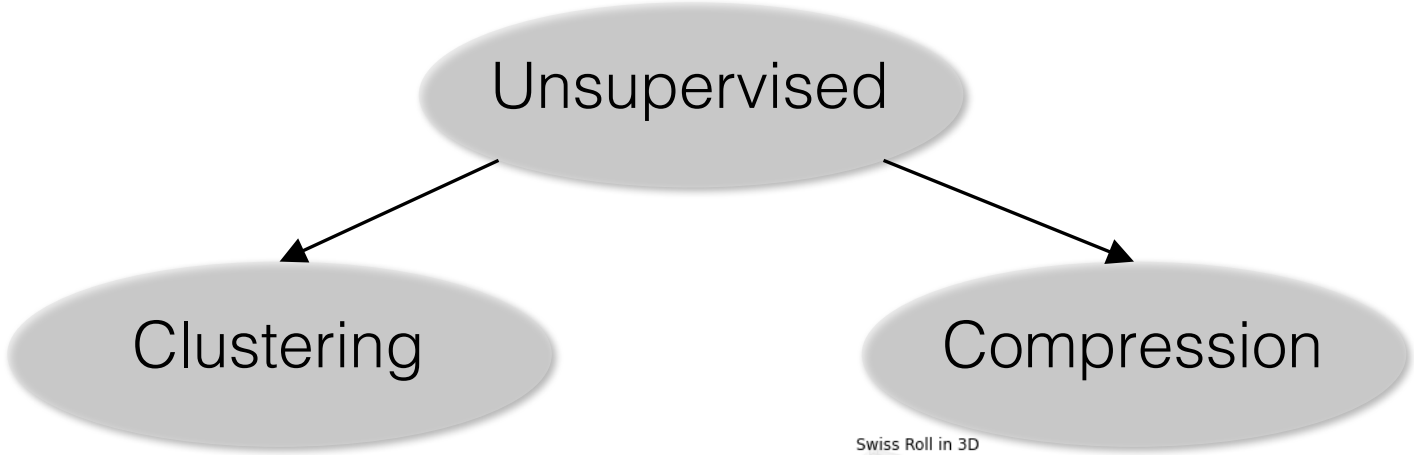
Stock prediction

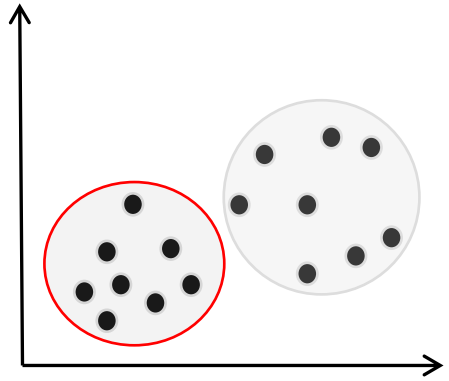
* Predict the price of a stock (y)
  * Depends on x =
    - Recent history of stock price
    - News events
    - Related commodities

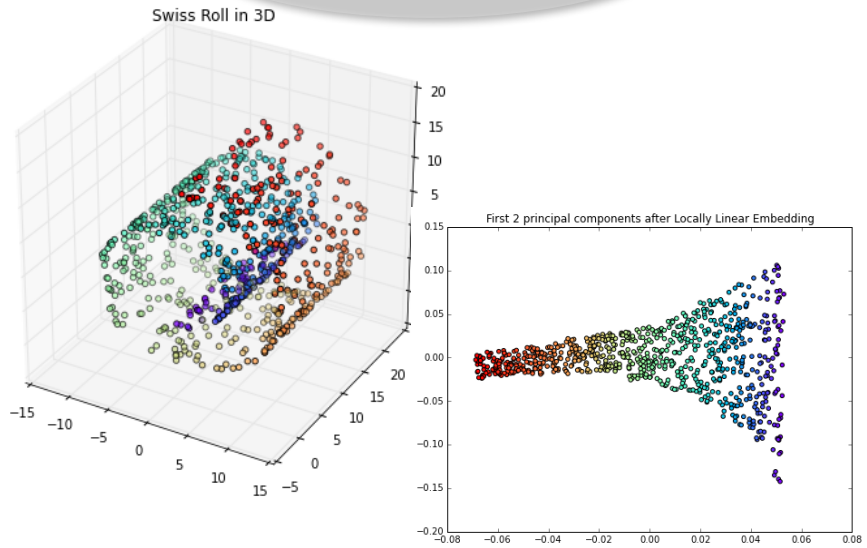
Spam or Not spam emails

Music or Tweeter Sentiment Analysis

Unsupervised Learning



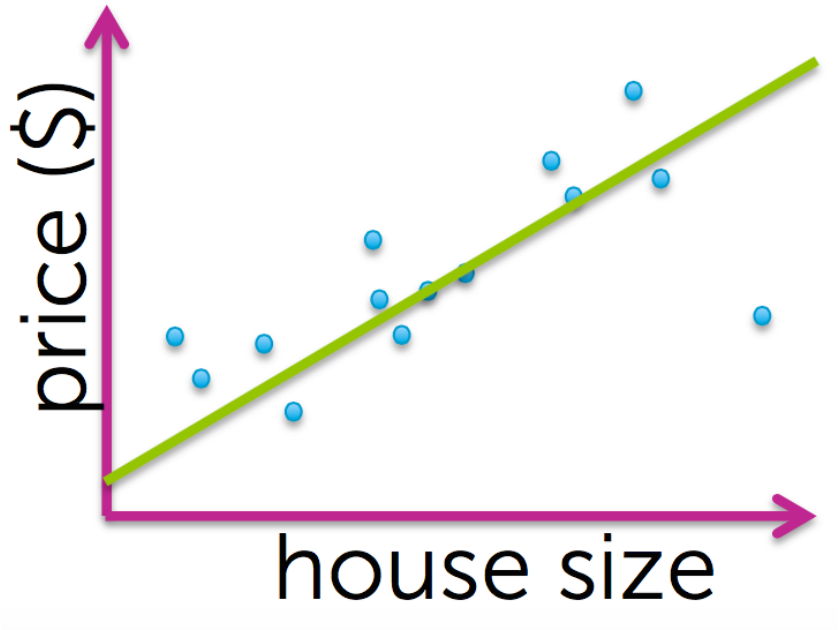




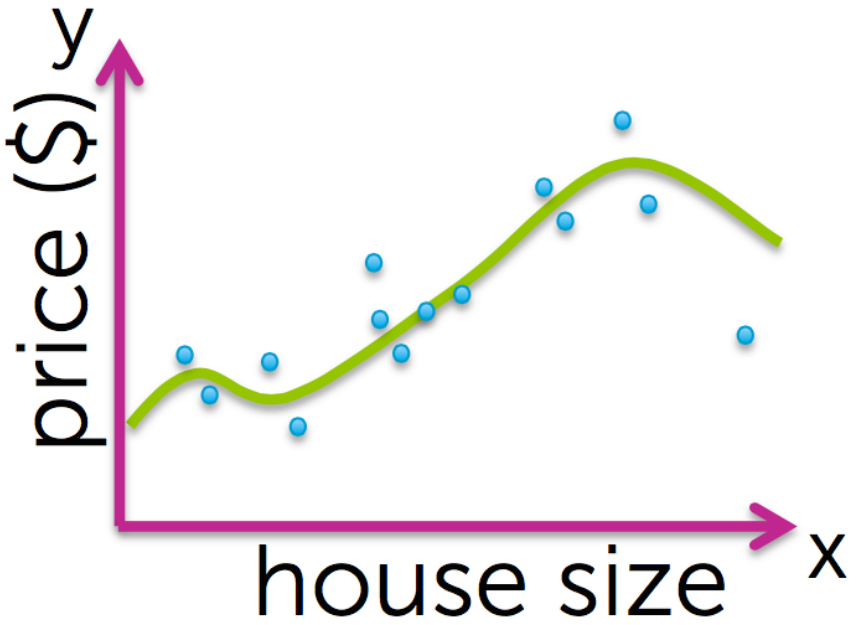
Simple Regression vs. Multiple Regression

What makes it simple?

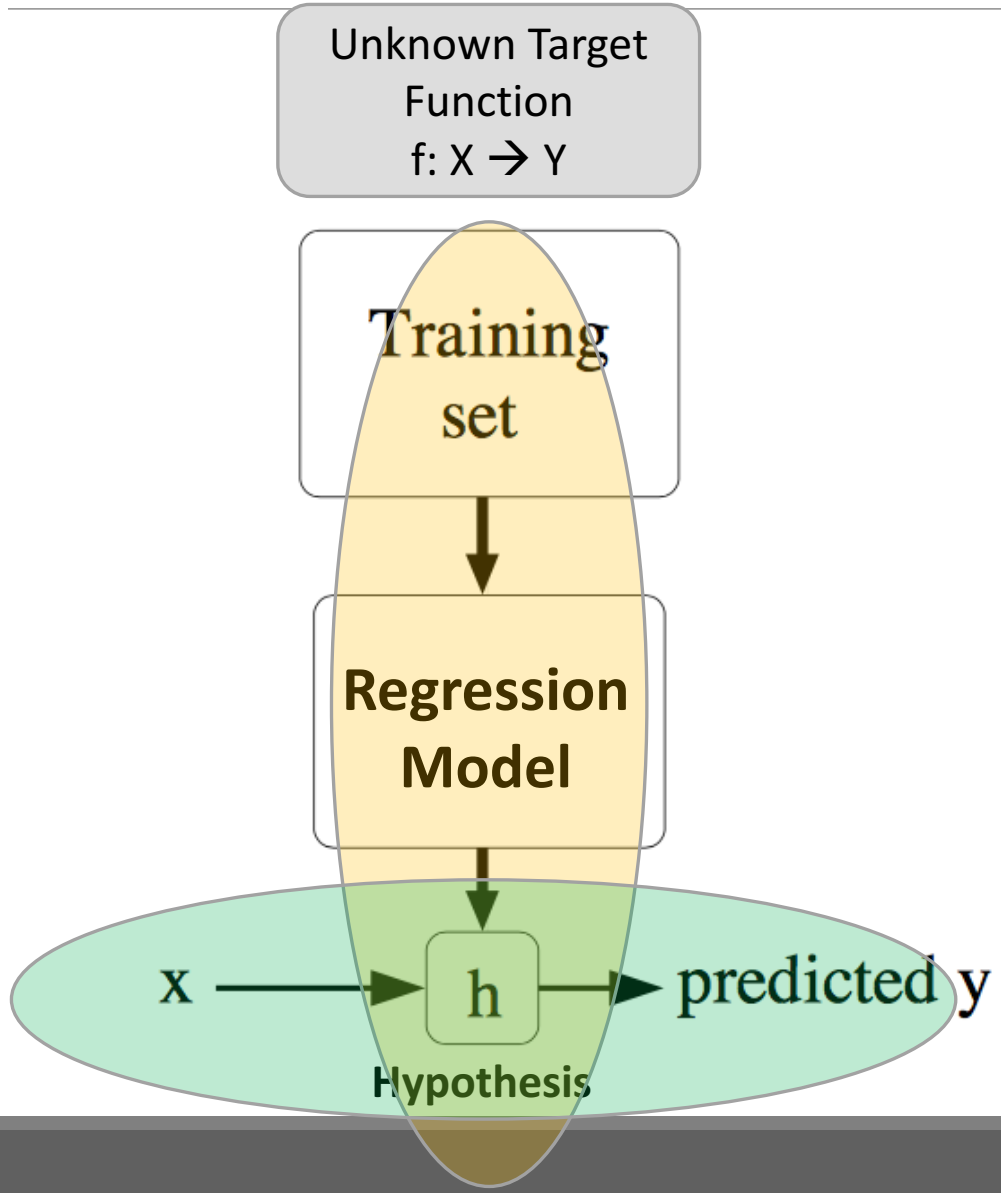
1 input and just fit a line to data

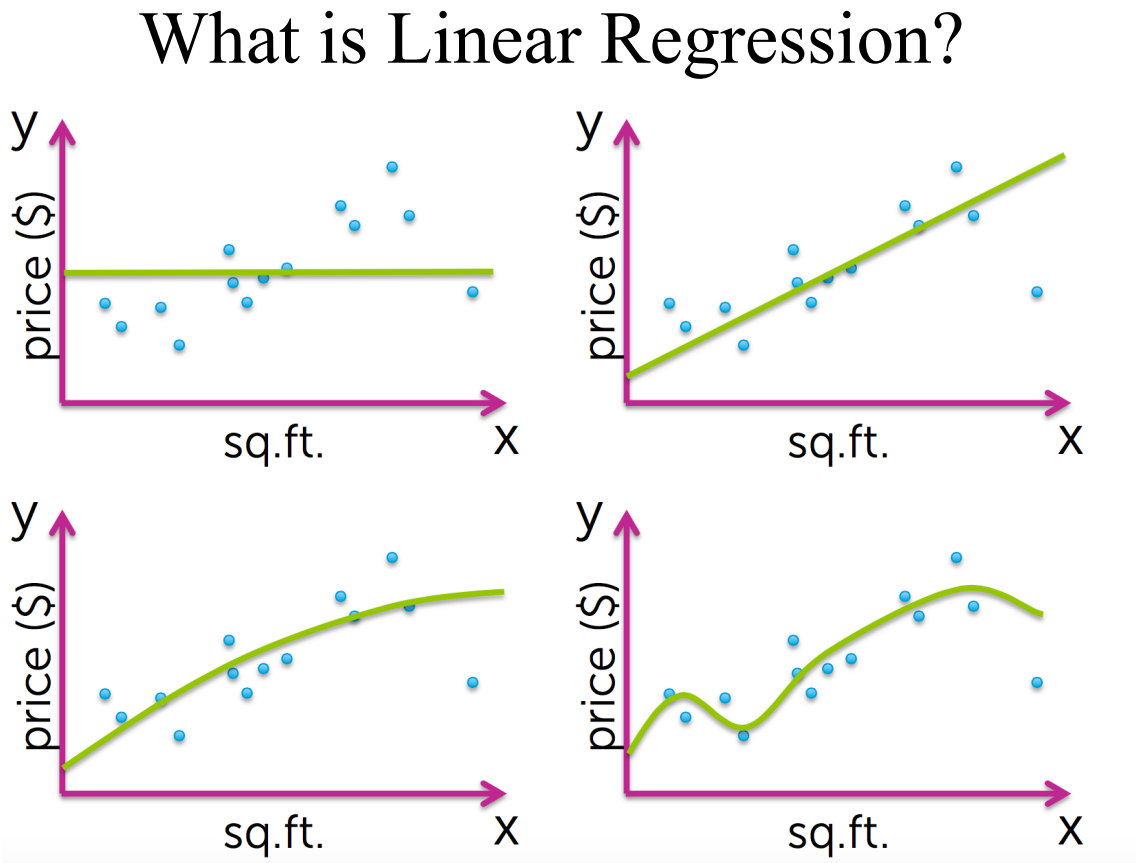


Fit more complex relationships than just a line

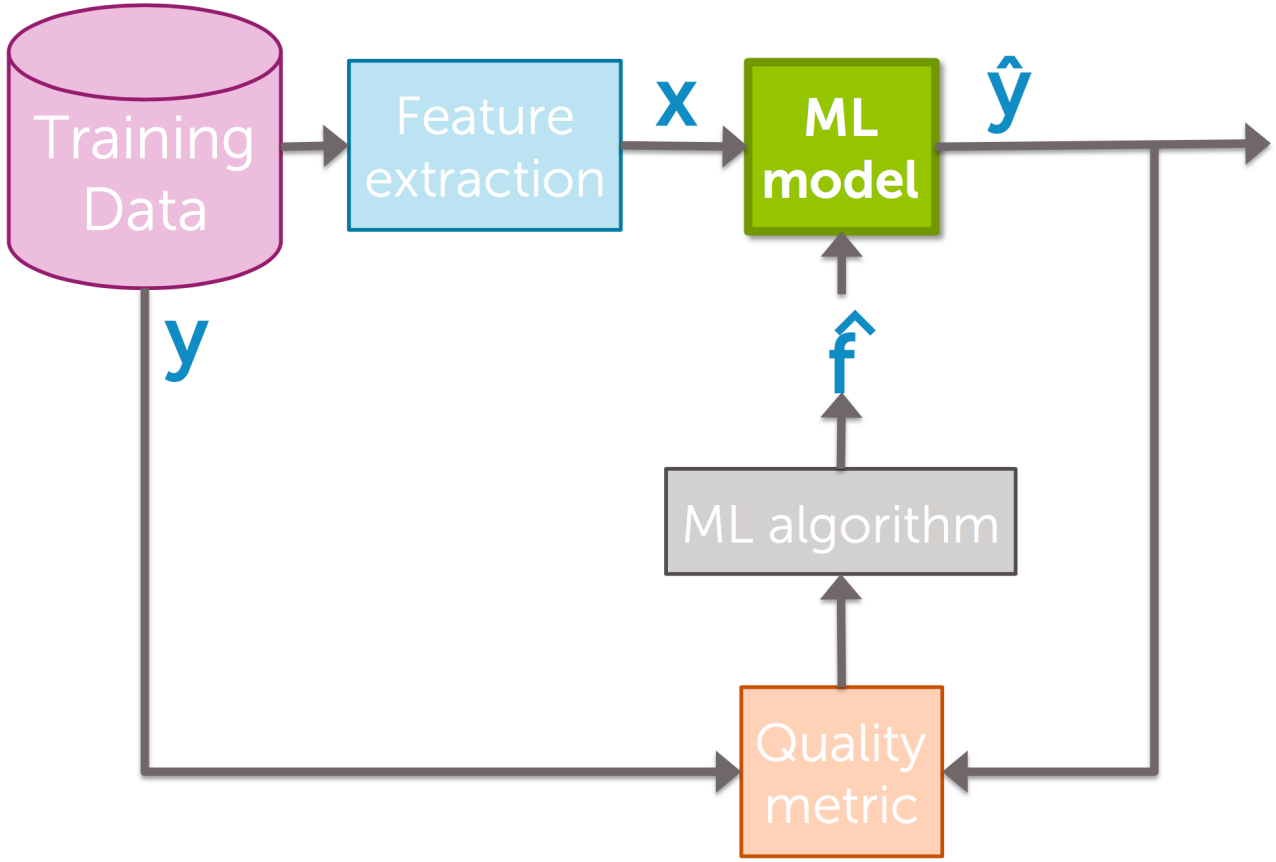


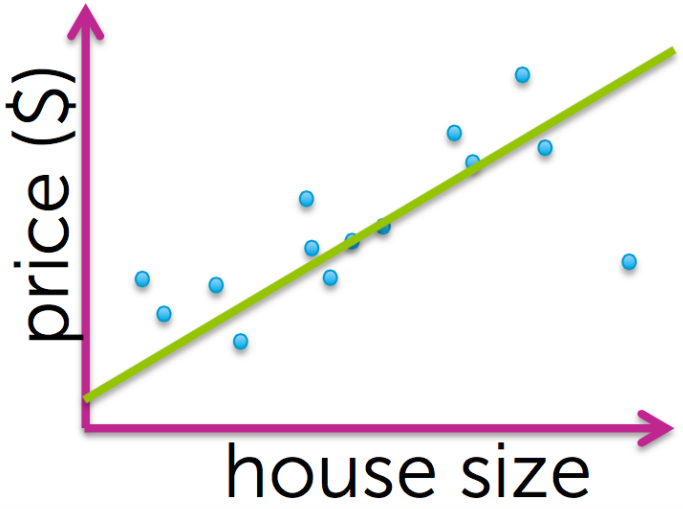
Linear Regression Learning Model





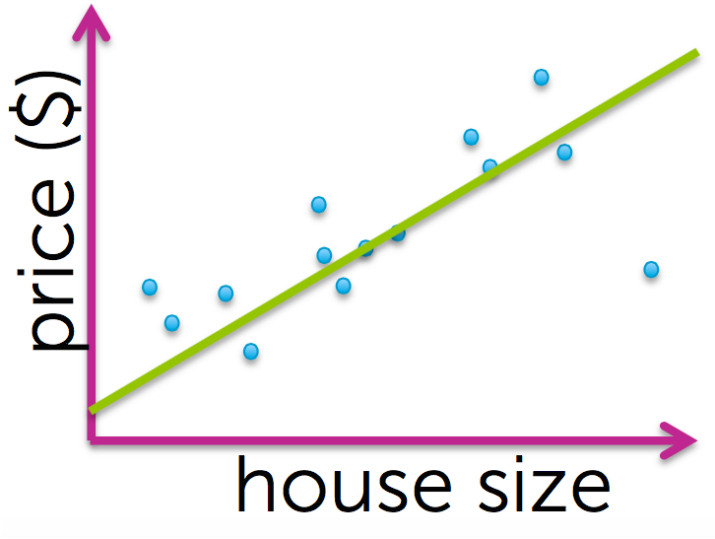
Regression Model





Predicting house prices

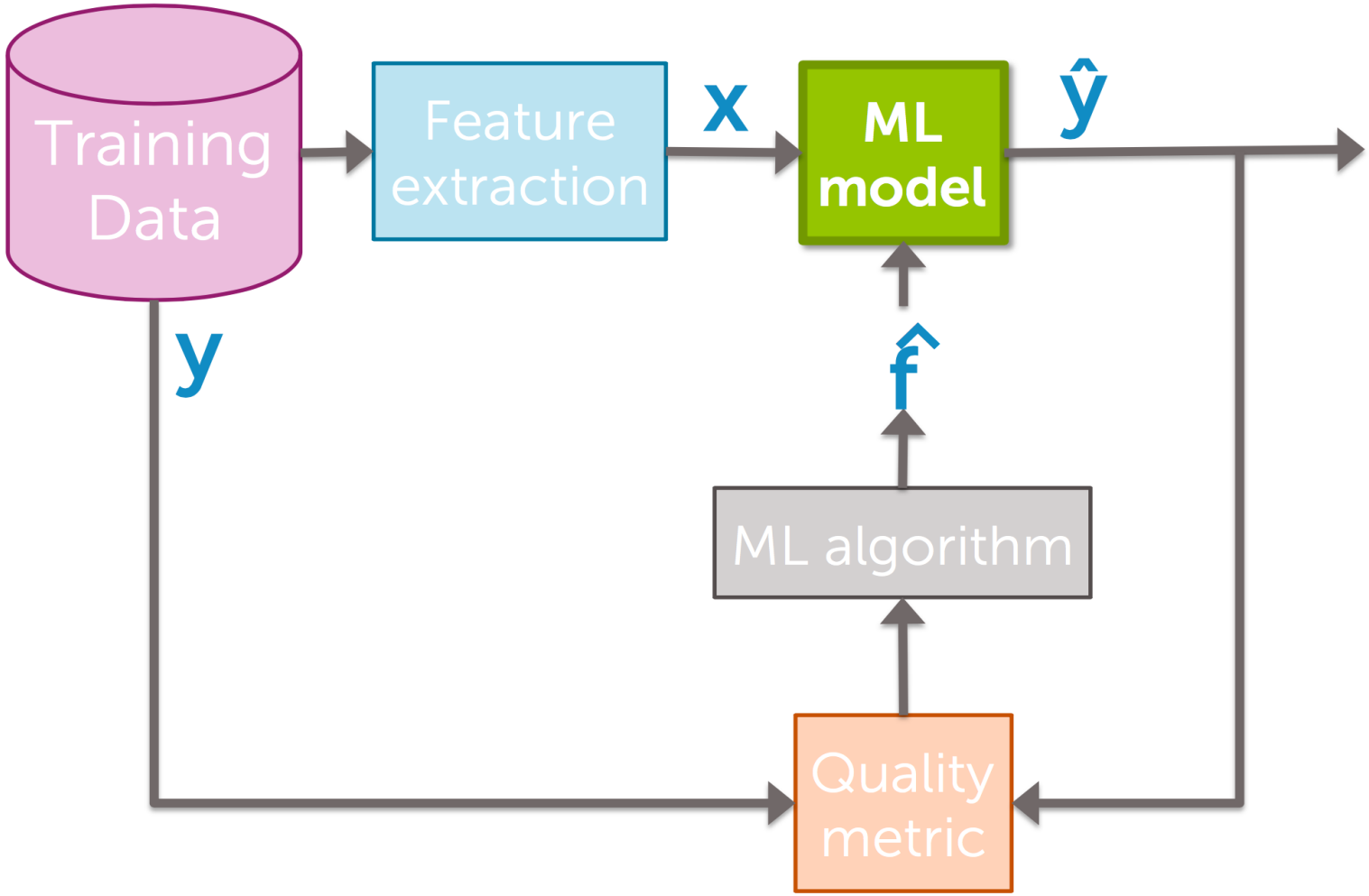
How much is my house worth?



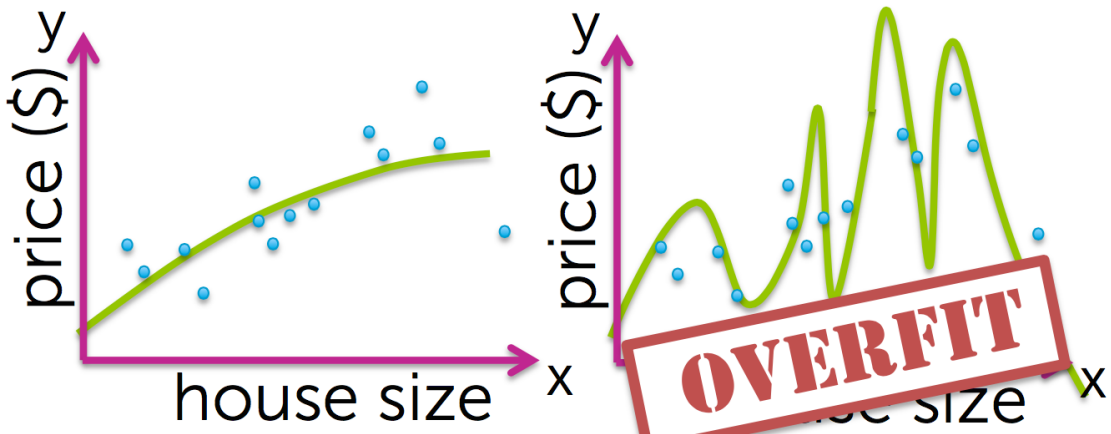
Look at recent sales in my neighborhood

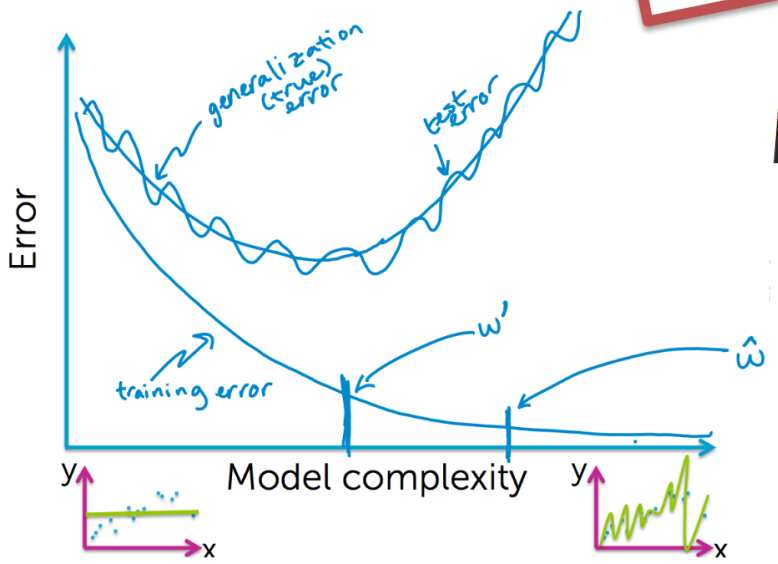
- How much did they sell for?

Regression Model



Assessing Performance





Measures of error:

- Training
- Test
- True (generalization)