

# Machine Learning

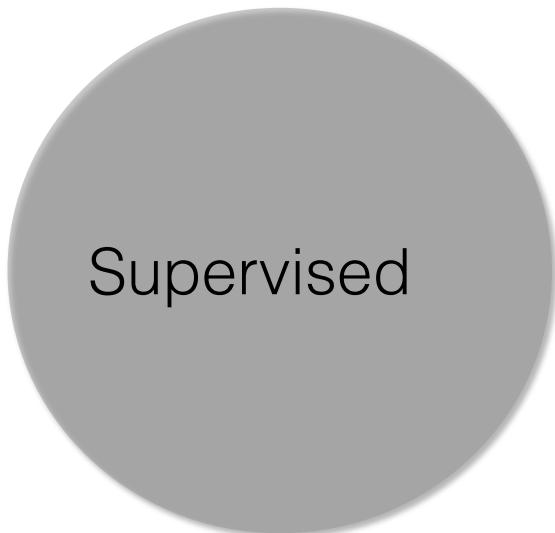


Paul Rad, Ph.D.

Associate Professor  
Cyber Analytics and AI  
Information Systems and Cyber Security  
College of Business School  
210.872.7259

# 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

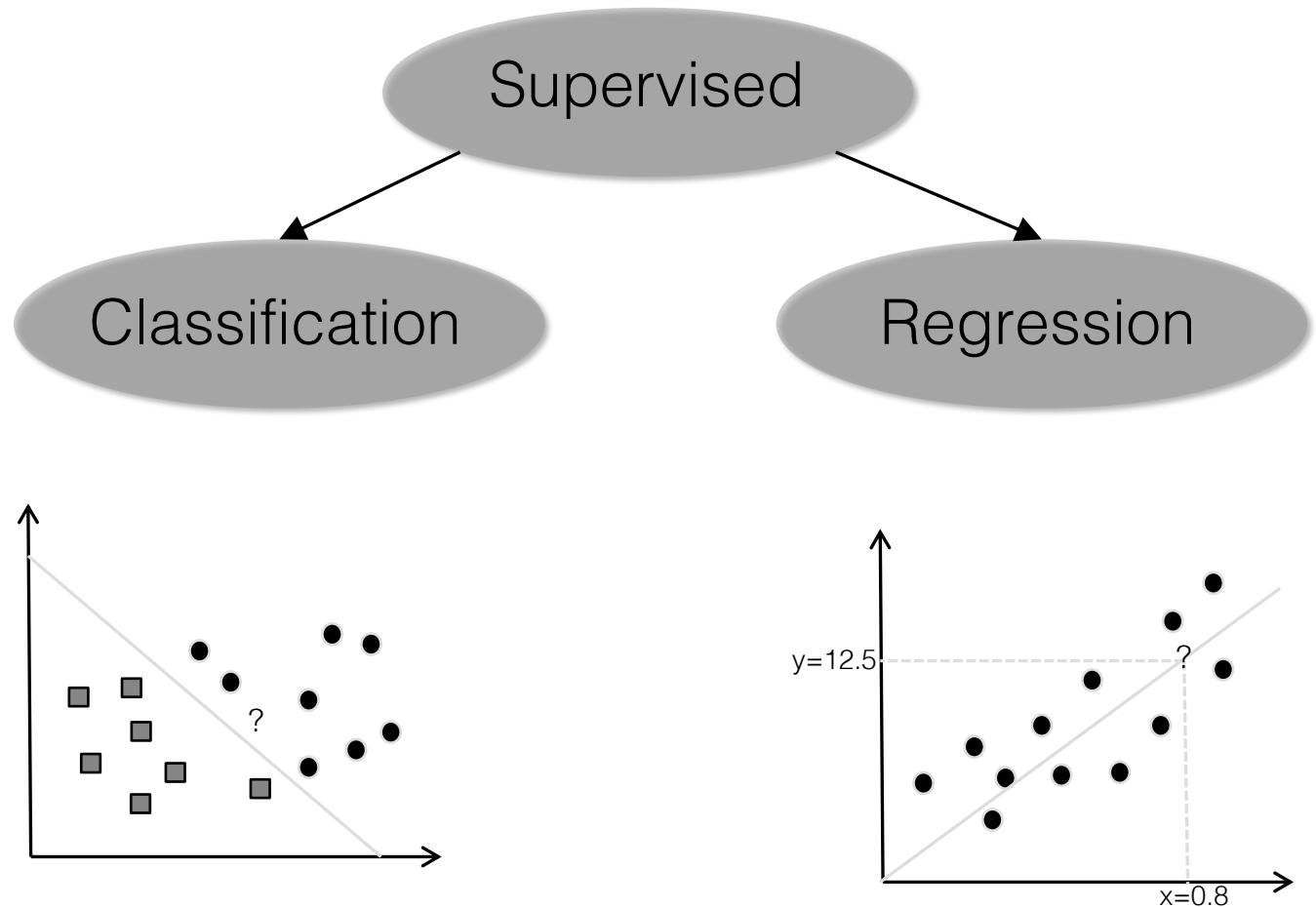
**Given** examples of a function  $(X, F(X))$

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

Discrete  $F(X)$ : Classification

Continuous  $F(X)$ : Regression

$F(X) = \text{Probability}(X)$ : Probability estimation



# Regression and Classification Examples

---

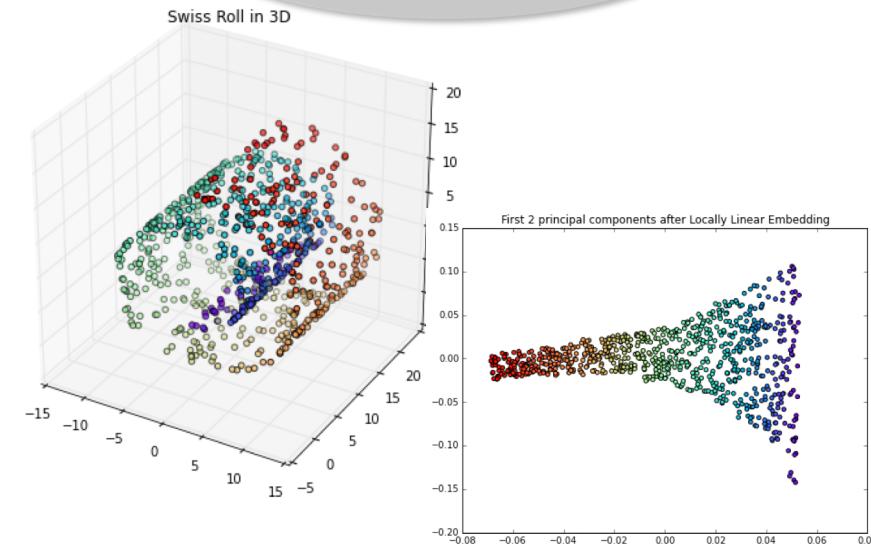
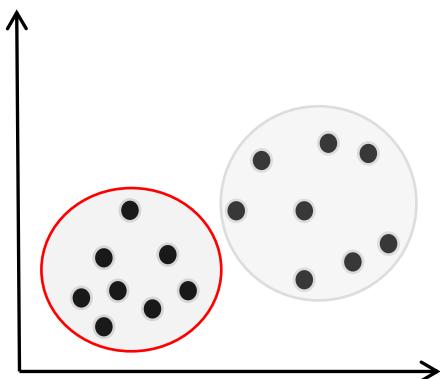
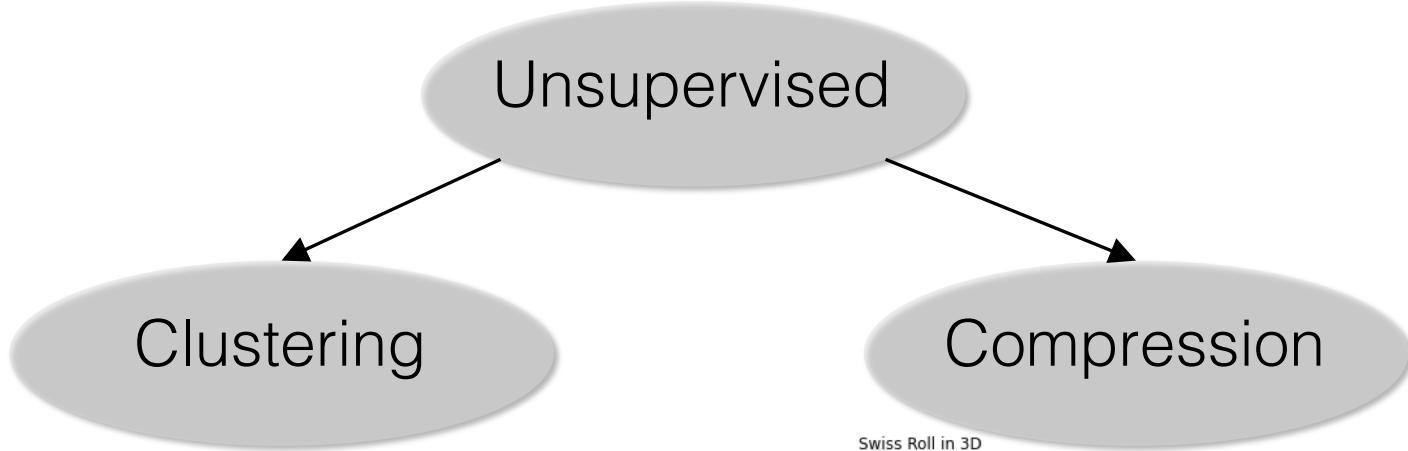
## Stock prediction

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

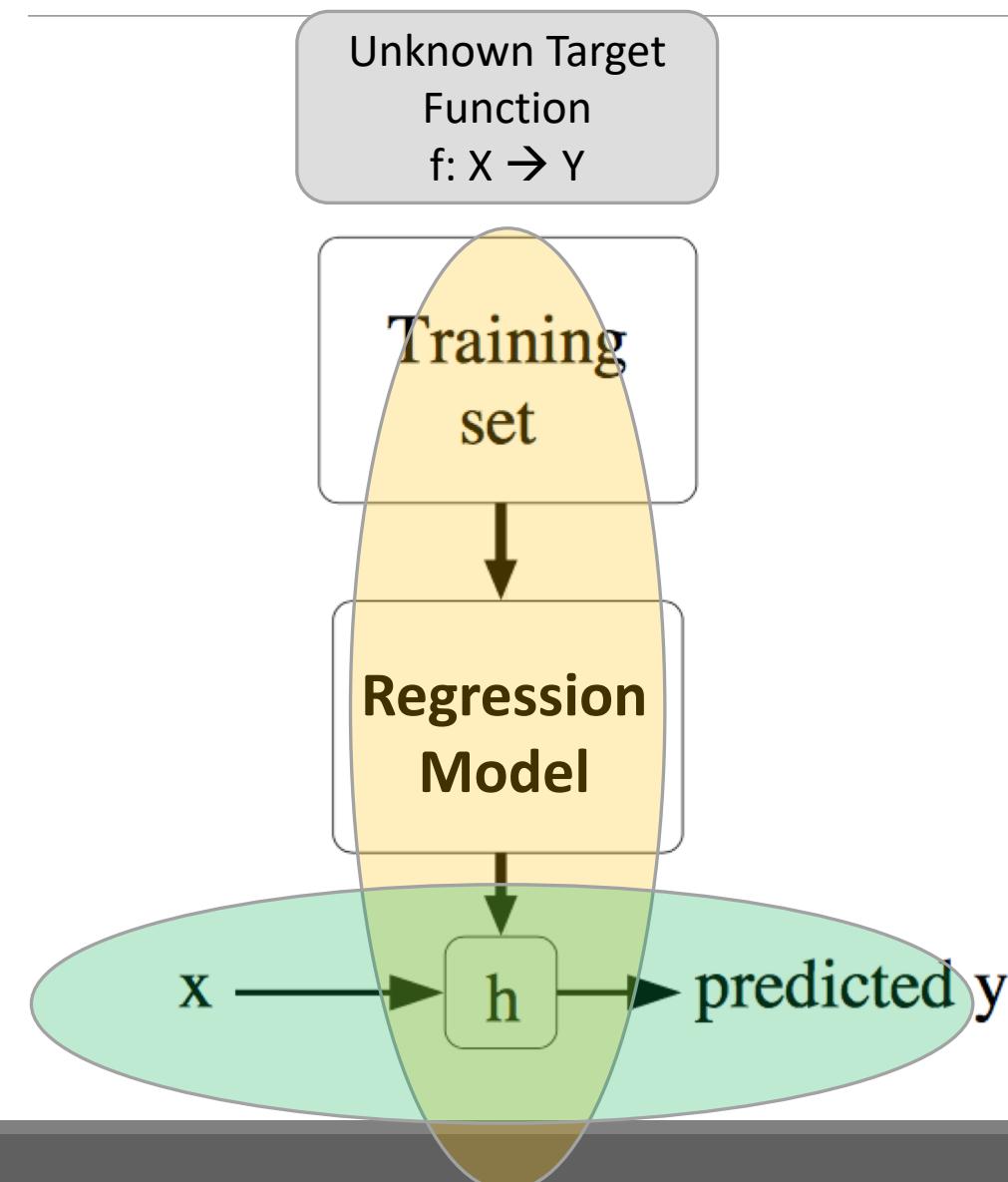
Spam or Not spam emails

Music or Tweeter  
Sentiment Analysis

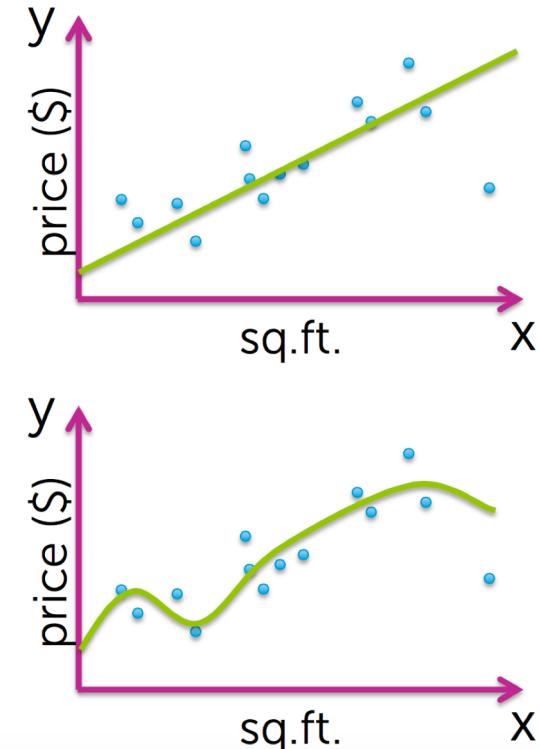
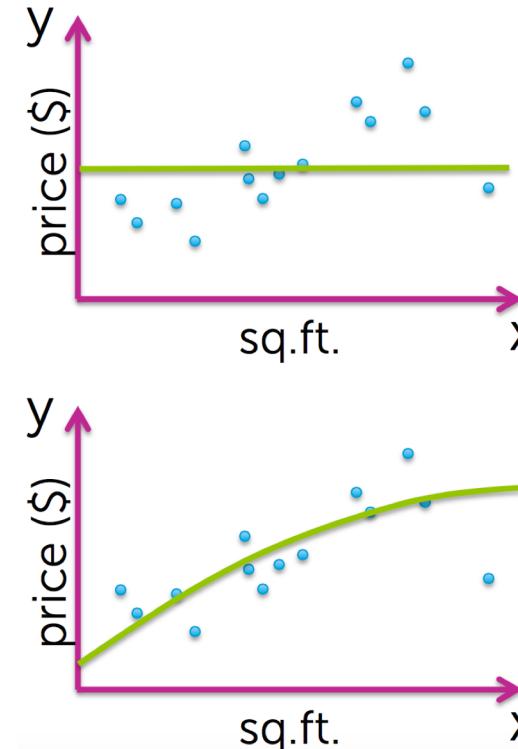
# Unsupervised Learning



# Linear Regression Learning Model



What is Linear Regression?

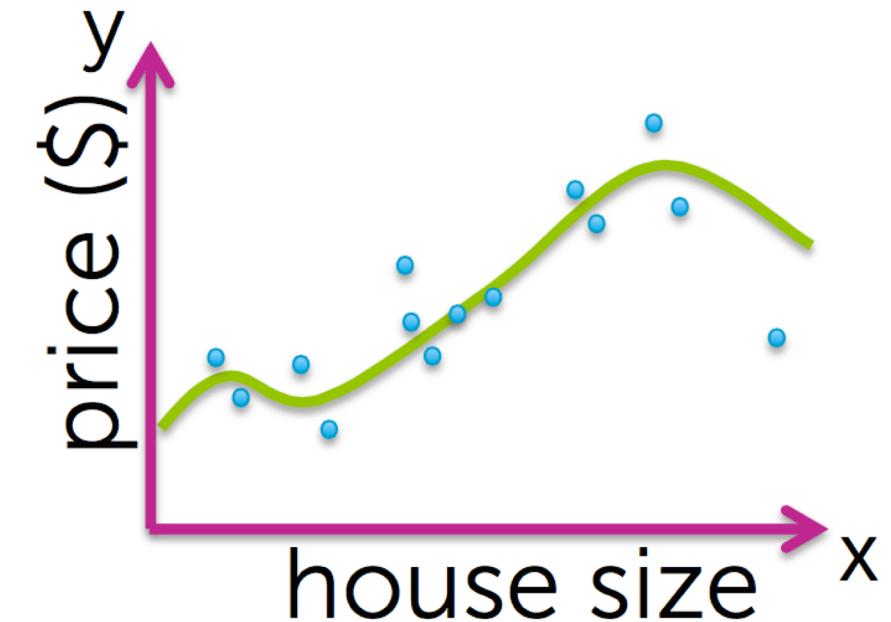
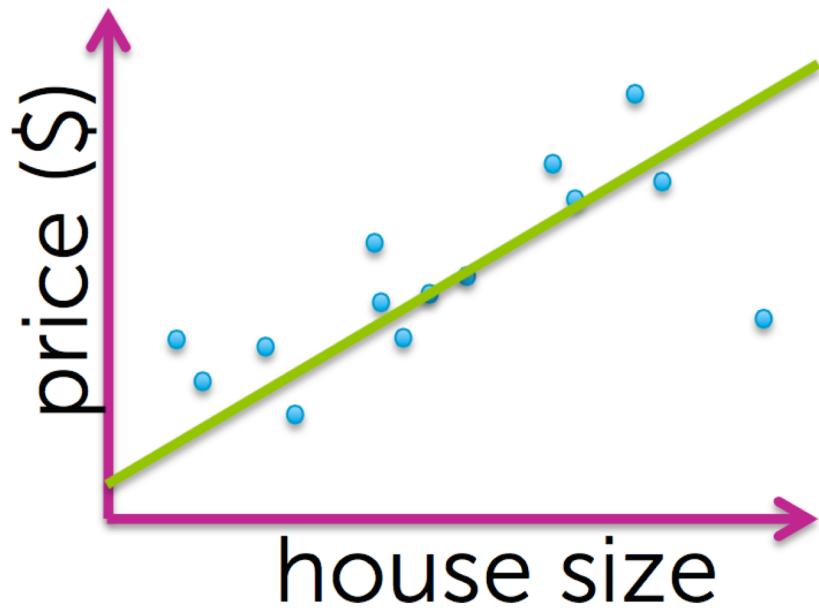


# Simple Regression vs. Multiple Regression

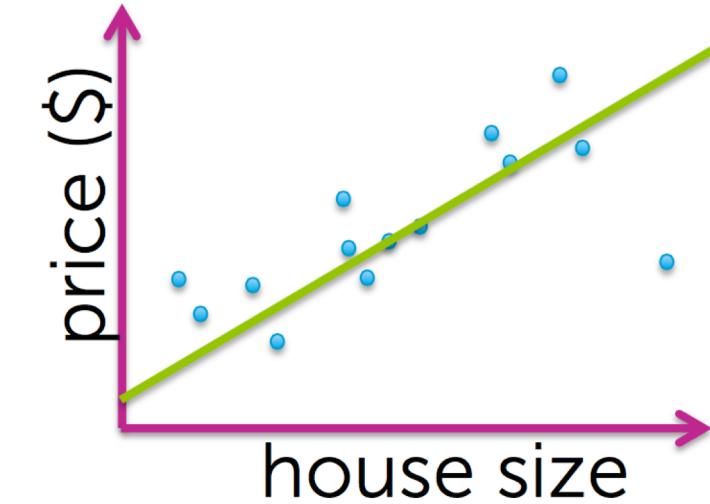
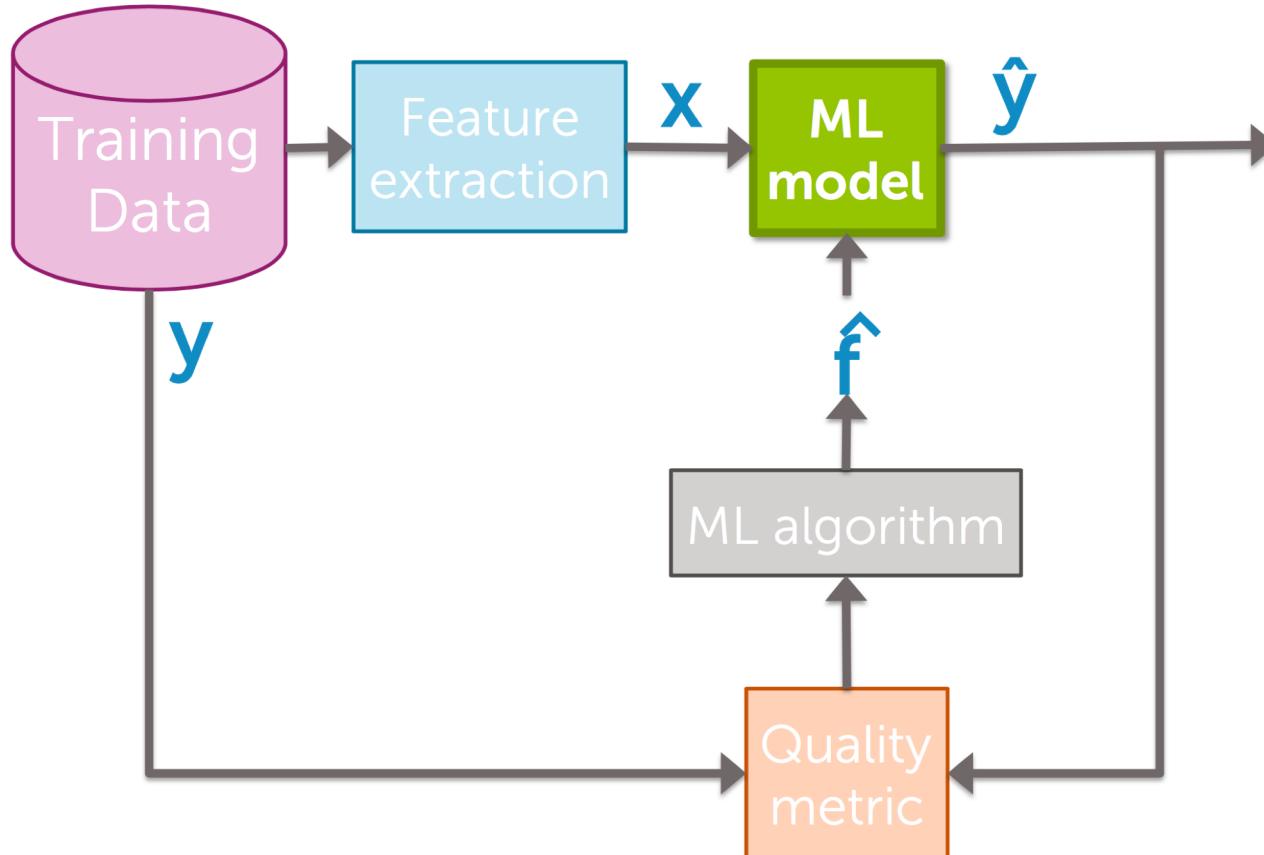
What makes it simple?

1 input and just fit a line to data

Fit **more complex**  
relationships than just a line



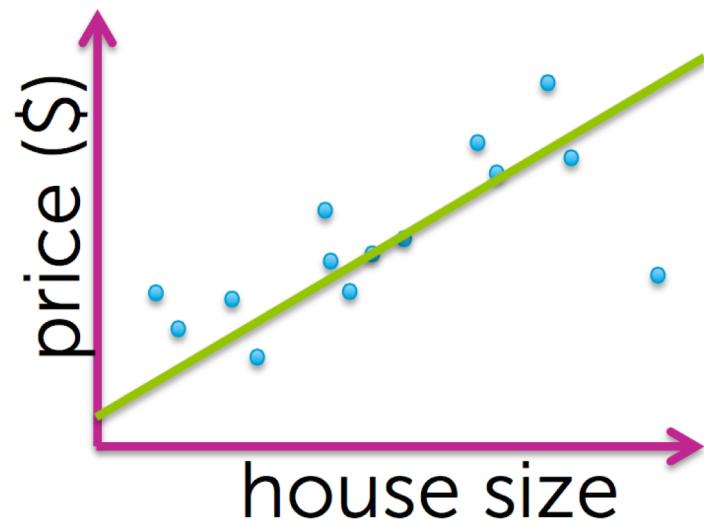
# Regression Model



# Predicting house prices

---

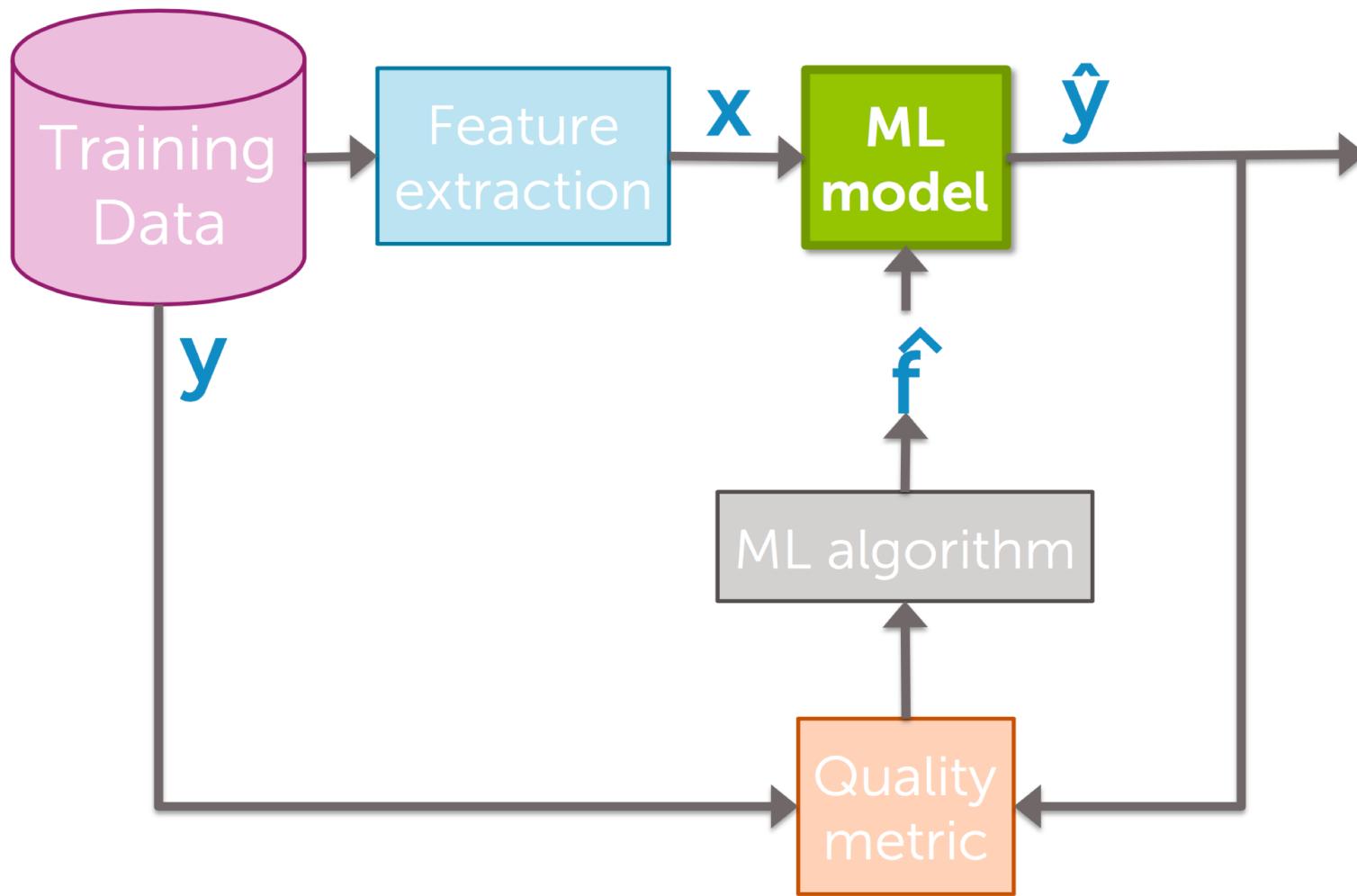
How much is my house worth?



Look at recent sales in my neighborhood

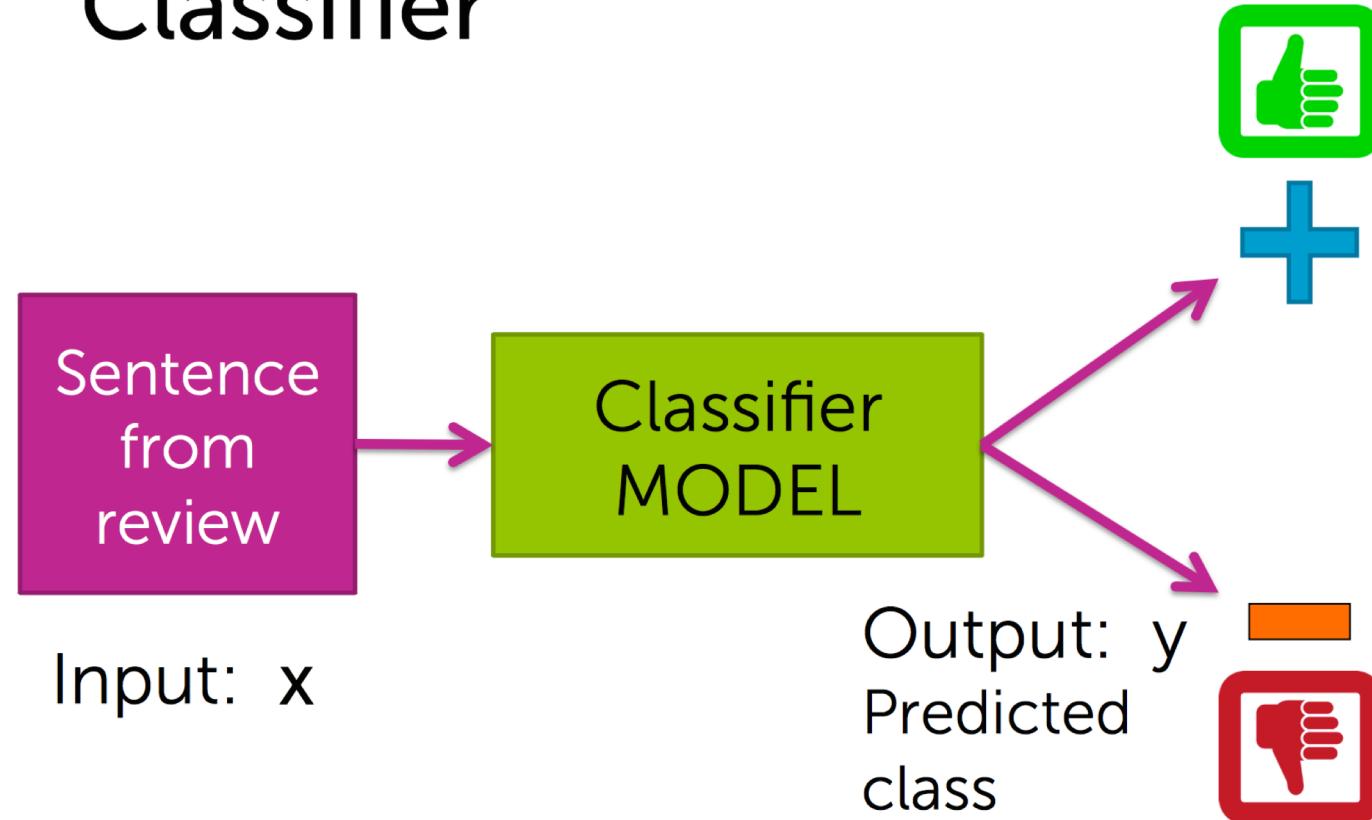
- How much did they sell for?

# Regression Model



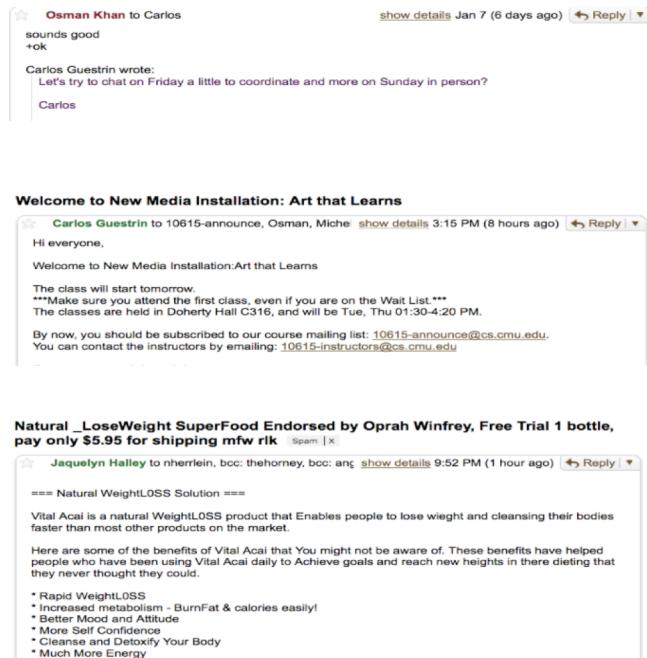
# Sentiment Analysis or Net Promoter Score (NPS)

## Classifier

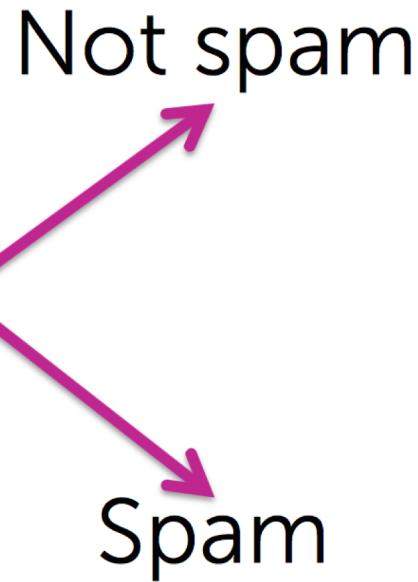


# Spam Filtering

## Spam filtering



Input: x



Output: y

# Sentiment Classifier

Will use training data to learn a weight or coefficient for each word

| Word                     | Coefficient                              | Value |
|--------------------------|--|-------|
|                          | $\hat{w}_0$                              | -2.0  |
| good                     | $\hat{w}_1$                              | 1.0   |
| great                    | $\hat{w}_2$                              | 1.5   |
| awesome                  | $\hat{w}_3$                              | 2.7   |
| bad                      | $\hat{w}_4$                              | -1.0  |
| terrible                 | $\hat{w}_5$                              | -2.1  |
| awful                    | $\hat{w}_6$                              | -3.3  |
| restaurant, the, we, ... | $\hat{w}_7, \hat{w}_8, \hat{w}_9, \dots$ | 0.0   |
| ...                      |  | ...   |

# Training Data

---

Training data:  
N observations ( $\mathbf{x}_i, y_i$ )

| $x[1] = \#awesome$ | $x[2] = \#awful$ | $y = \text{sentiment}$ |
|--------------------|------------------|------------------------|
| 2                  | 1                | +1                     |
| 0                  | 2                | -1                     |
| 3                  | 3                | -1                     |
| 4                  | 1                | +1                     |
| 1                  | 1                | +1                     |
| 2                  | 4                | -1                     |
| 0                  | 3                | -1                     |
| 0                  | 1                | -1                     |
| 2                  | 1                | +1                     |



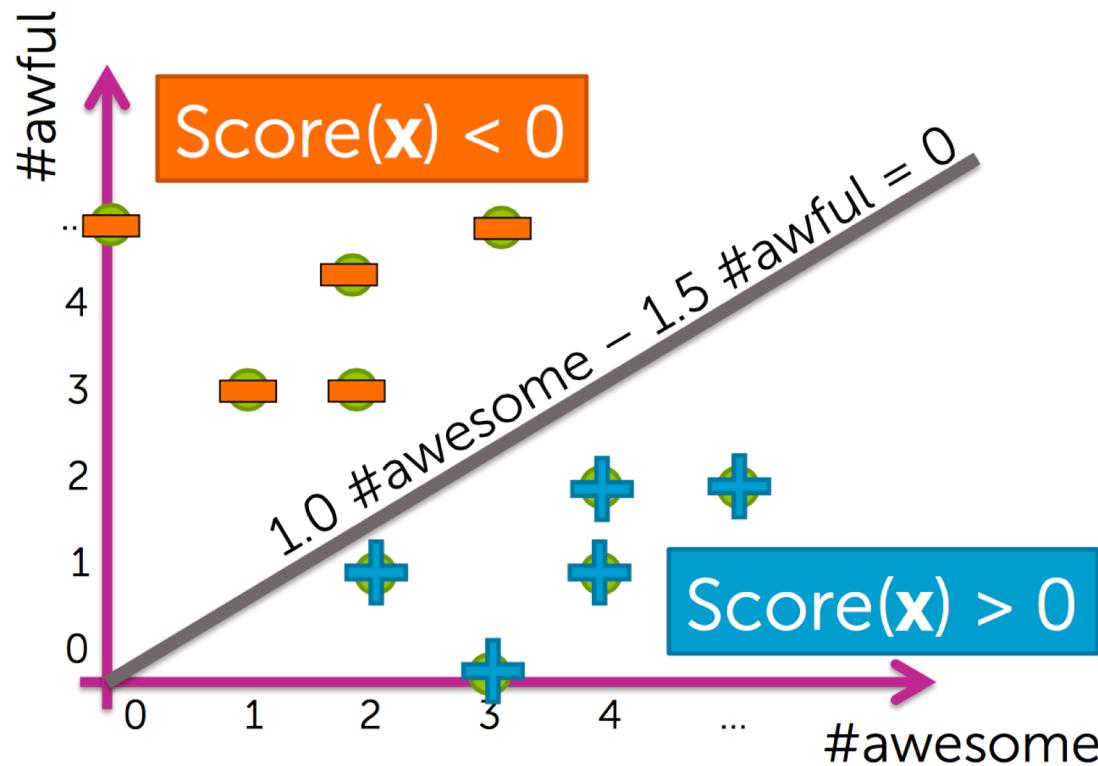
$\hat{\mathbf{w}}$

# Simple Model

| Word     | Coefficient |
|----------|-------------|
| #awesome | 1.0         |
| #awful   | -1.5        |

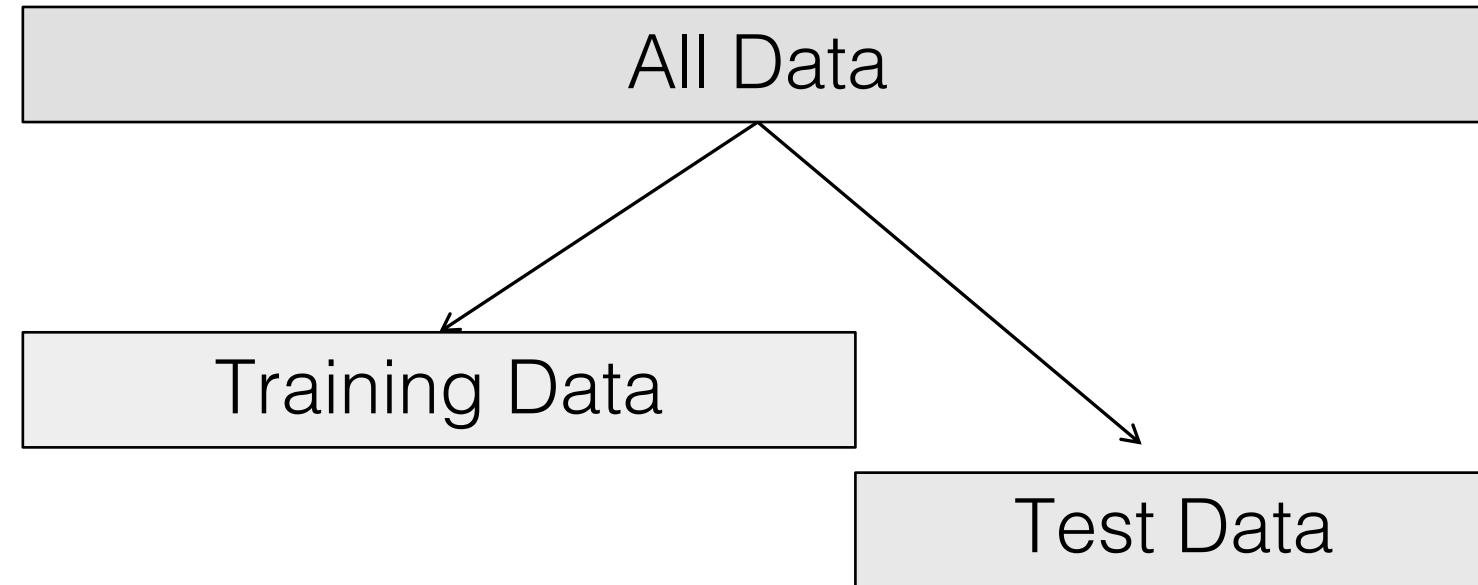


$$\text{Score}(x) = 1.0 \text{ #awesome} - 1.5 \text{ #awful}$$



# Training & Test Data

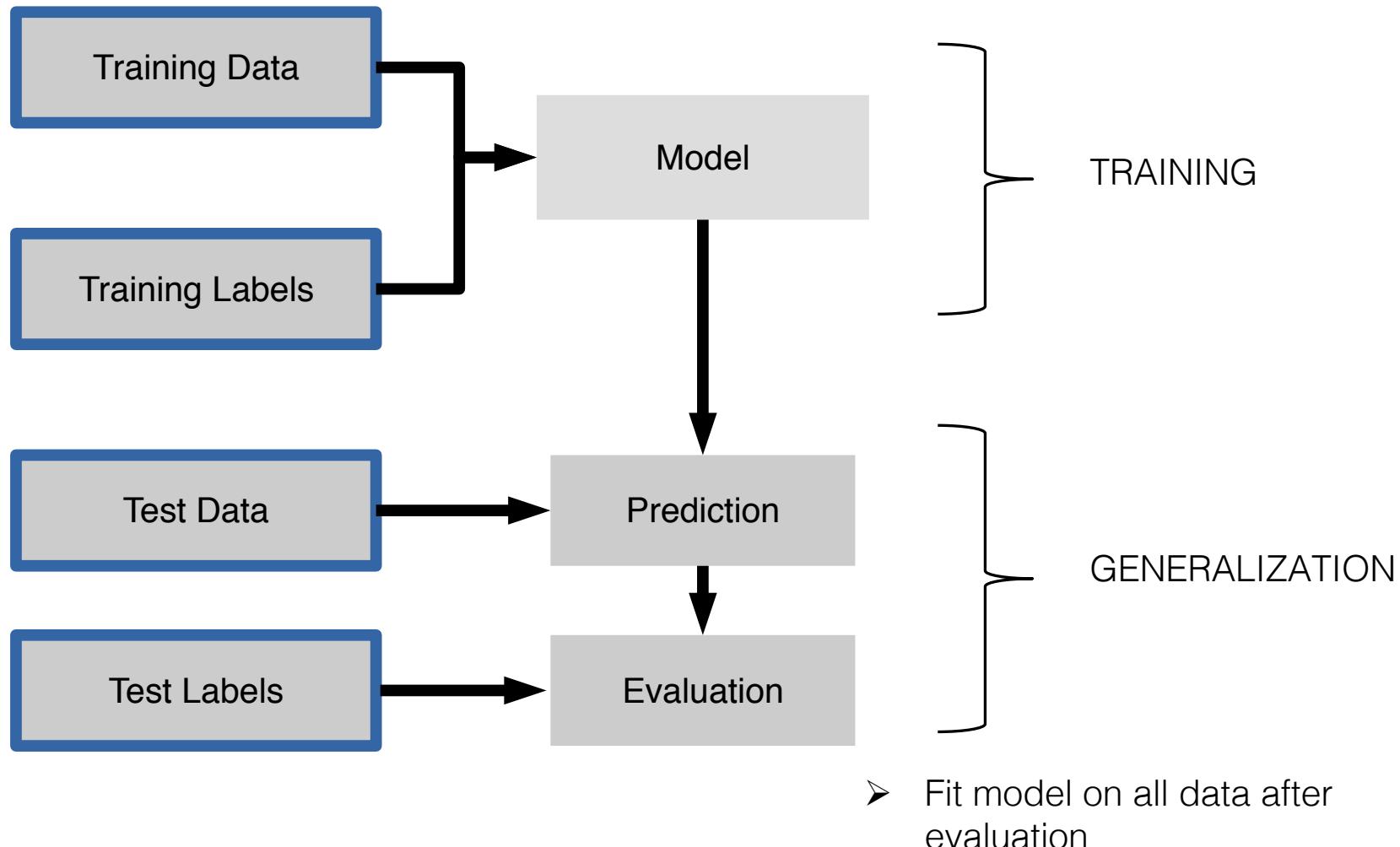
---



Typically:

- 75% : 25%
- 2/3 : 1/3

# Supervised Workflow



# Supervised Workflow

