Machine Learning

BY SONDRA IVERSON

Machine Learning

- The field of study that gives computers the ability to learn without being explicitly programmed. – Arthur Samuel

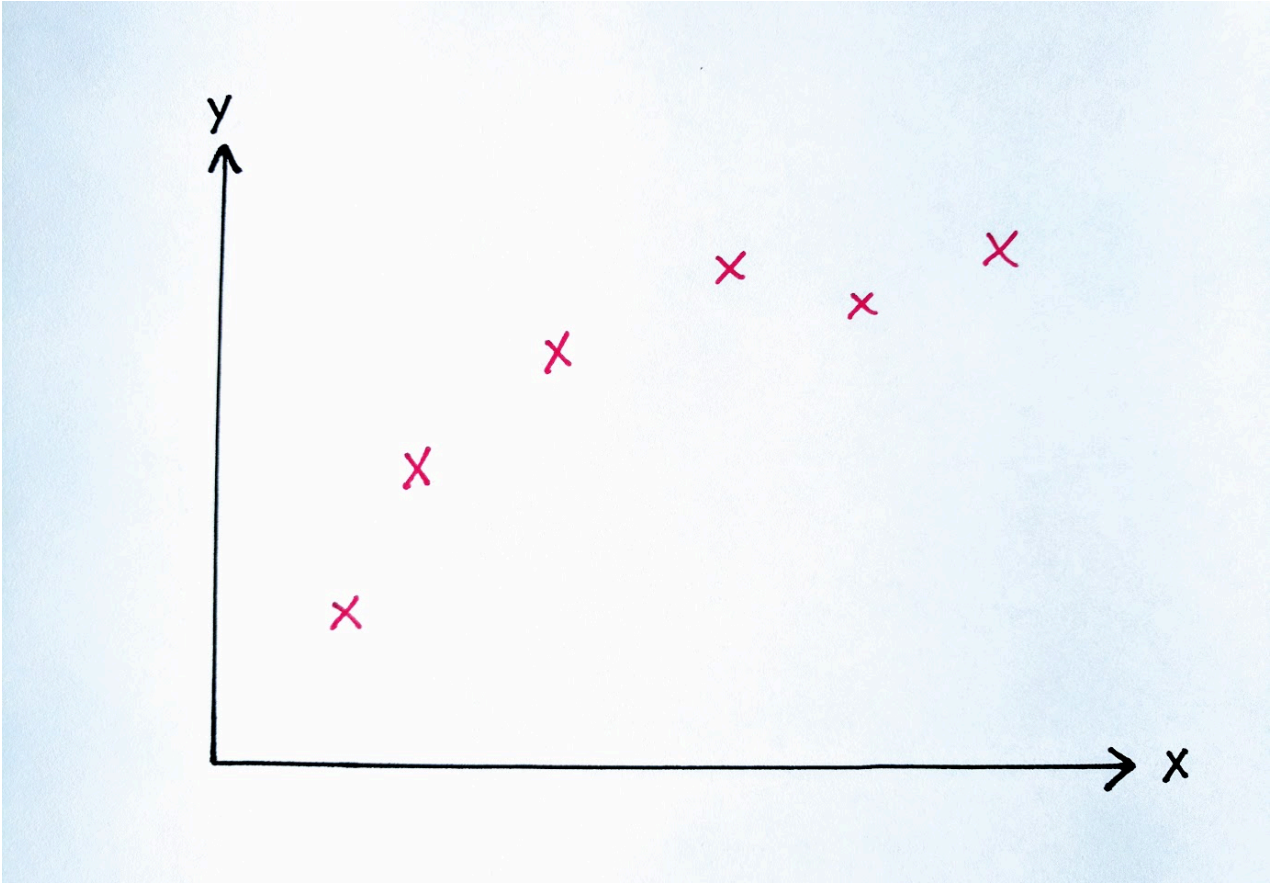
- Two Main Types of Algorithms
  - Supervised Learning – Given inputs and desired output, learn general rule
  - Unsupervised Learning – No labels given, try to find structure in data

Machine Learning Algorithms

- Supervised Learning
  - **Linear Regression**
  - Logistic Regression
  - Neural Networks
- Unsupervised Learning
  - Clustering
  - Anomaly Detection
  - Recommender Systems

Linear Regression

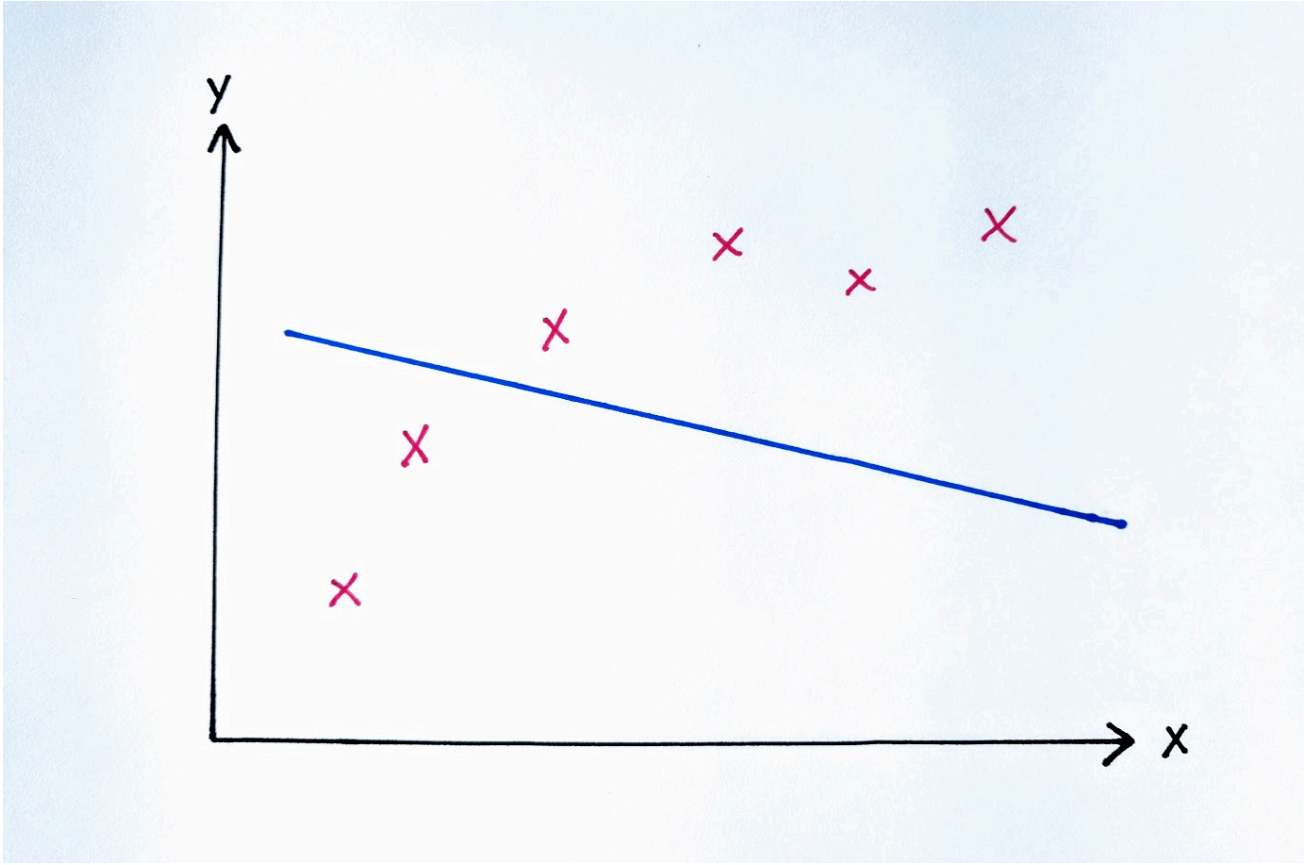
- For continuous data
- Example: predicting the selling price of a house based on the size
- x is called a feature, input, predictor or independent variable
- y is called a response, output, target or dependent variable



Linear Regression

$$h_\theta(x) = \theta_0 + \theta_1 x$$

- Hypothesis Function is the equation of a line

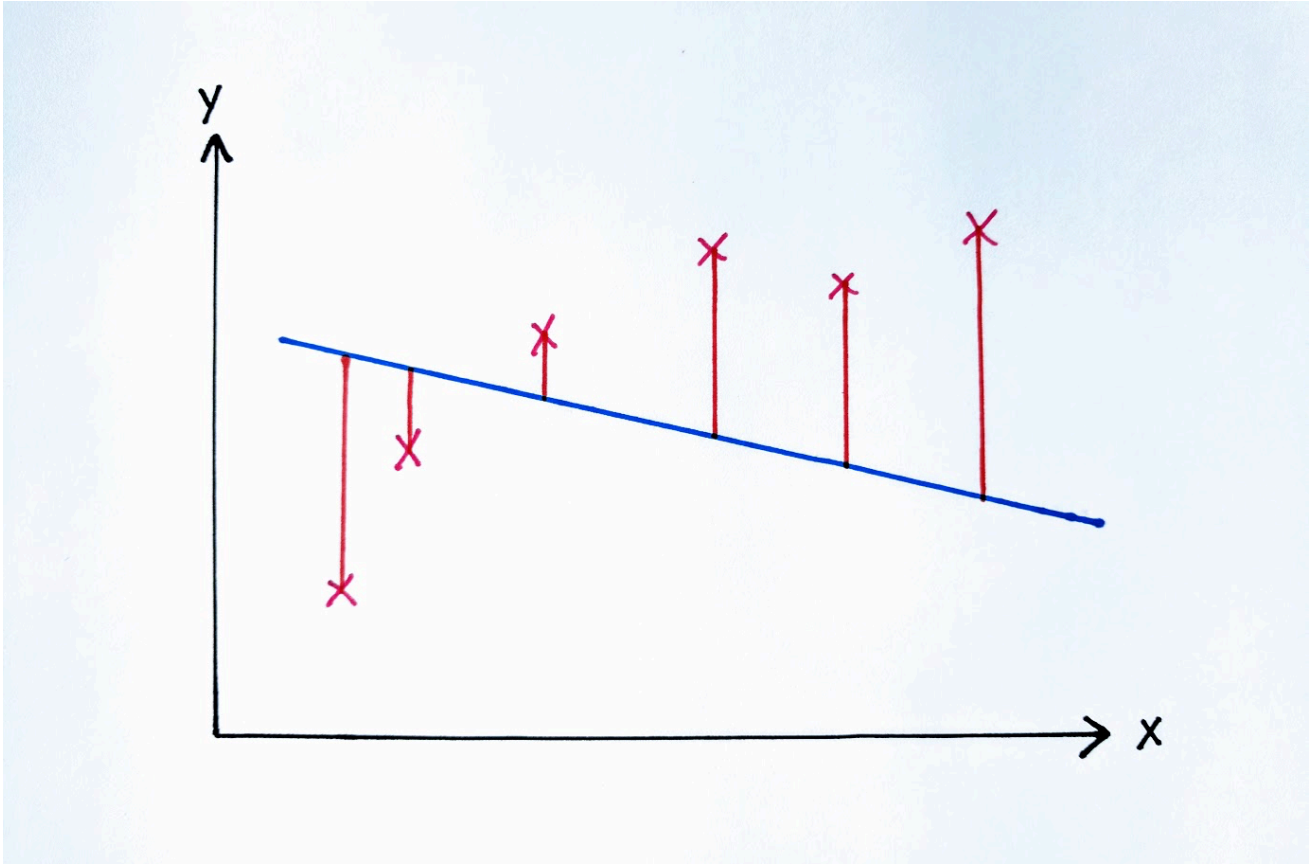


Linear Regression

- Calculate cost (i.e. error)

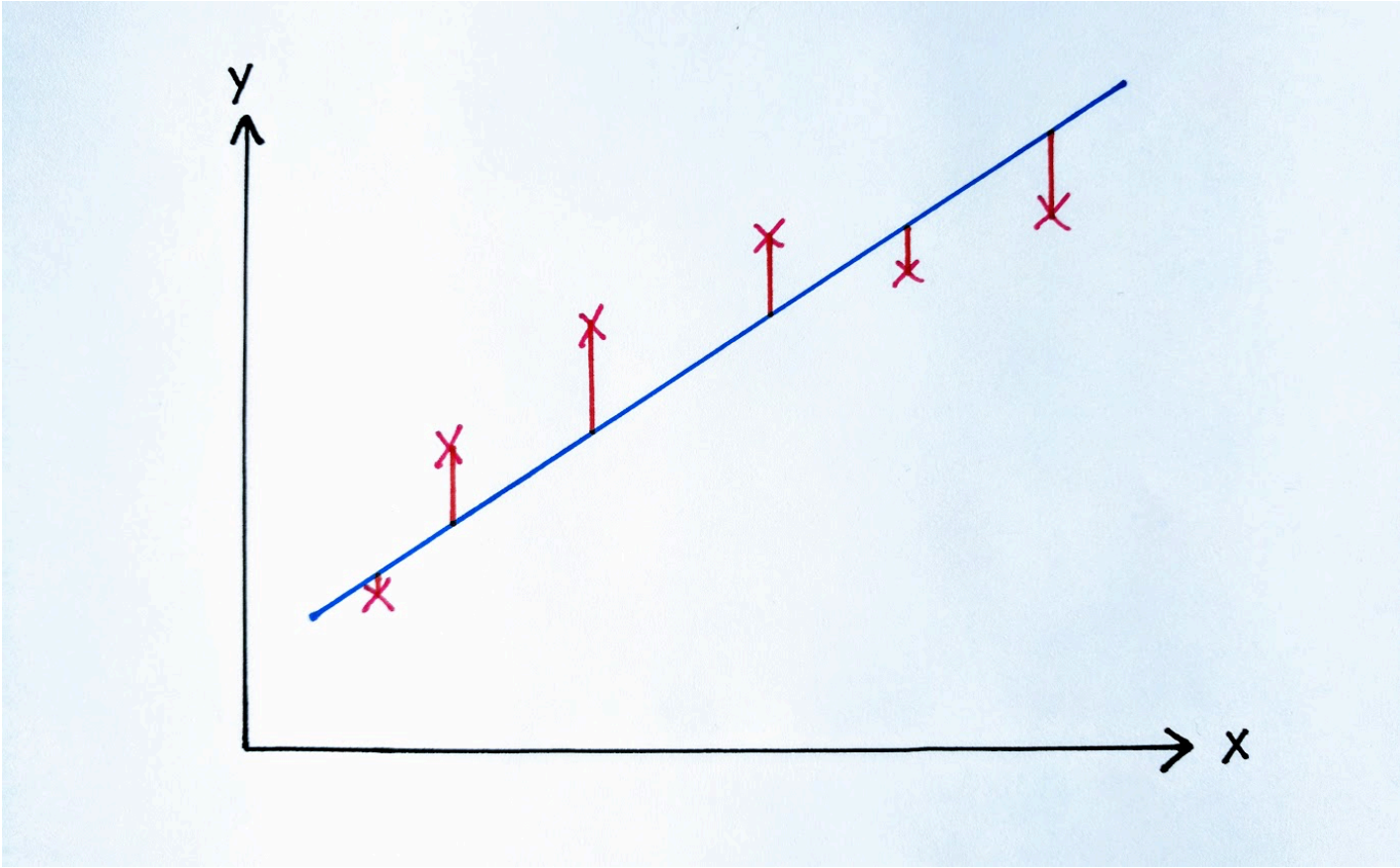
$$\sum_{i=1}^{m} \left(h_\theta(x_i) - y_i\right)^2$$

- Residual Sum of Squares (RSS)



Linear Regression

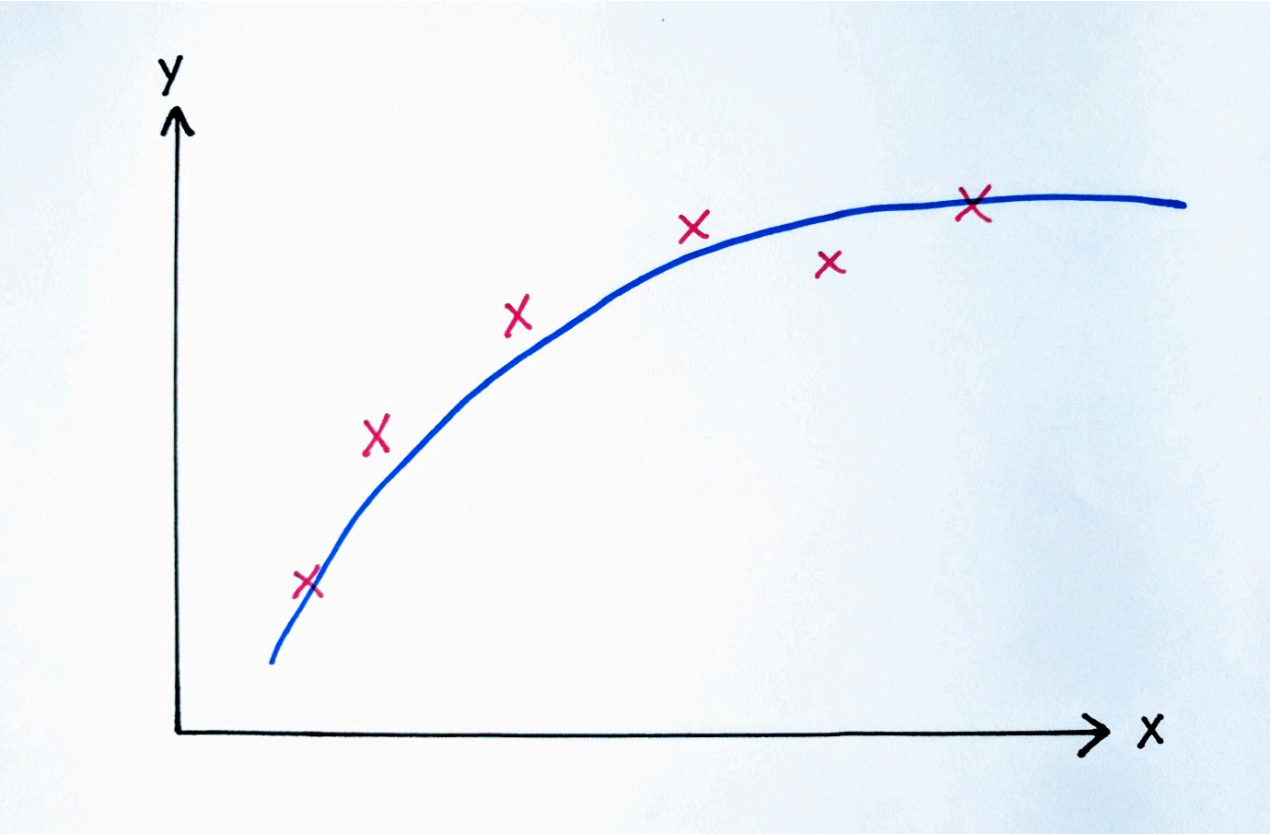
- Find the equation that minimizes the error for the training data
- This gives the theta values for the hypothesis function
- Use that equation to compute the output for new input values



Linear Regression

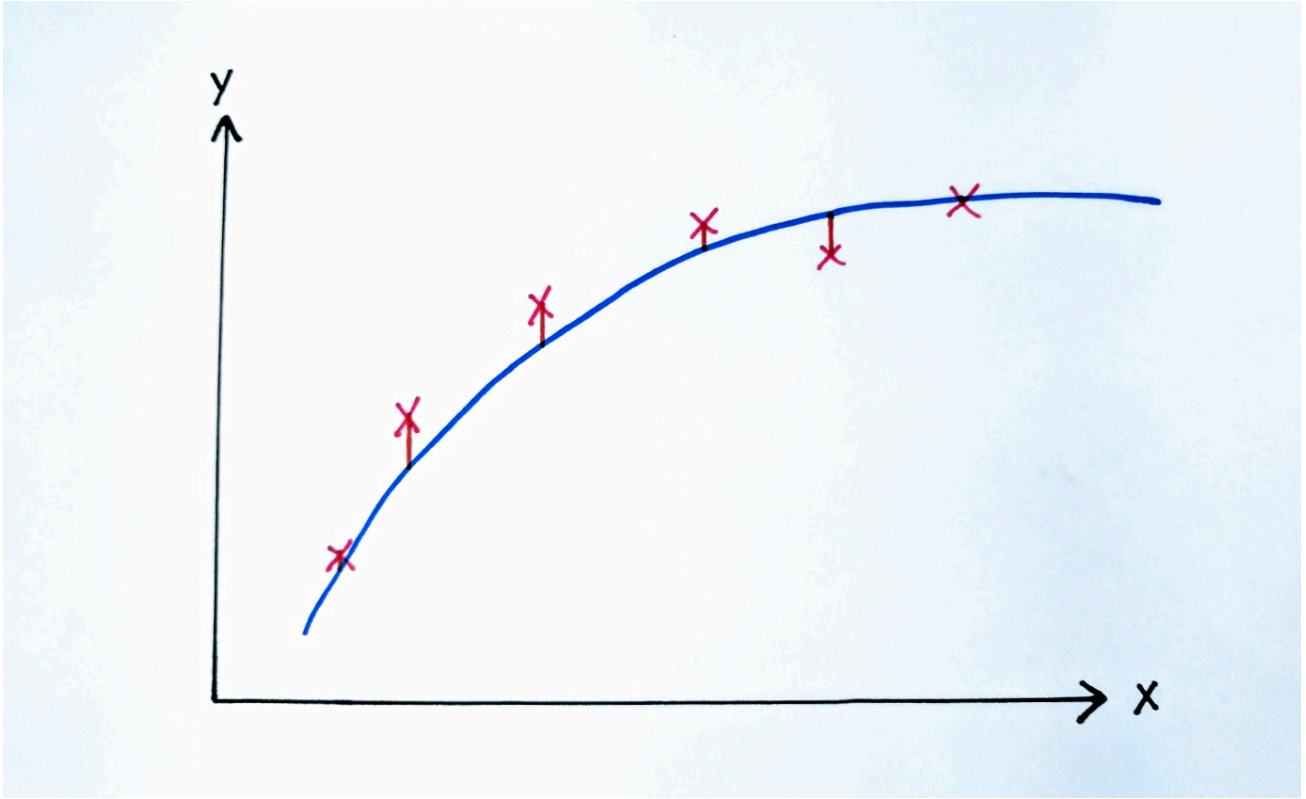
- Can use higher order terms for quadratic, cubic, etc. equations

$$h_\theta(x) = \theta_0 + \theta_1 x + \theta_2 x^2$$



Linear Regression

- Calculate and minimize cost with same algorithm
- Also works in multiple dimensions using multiple features
- Determines weights (positive or negative) for all features

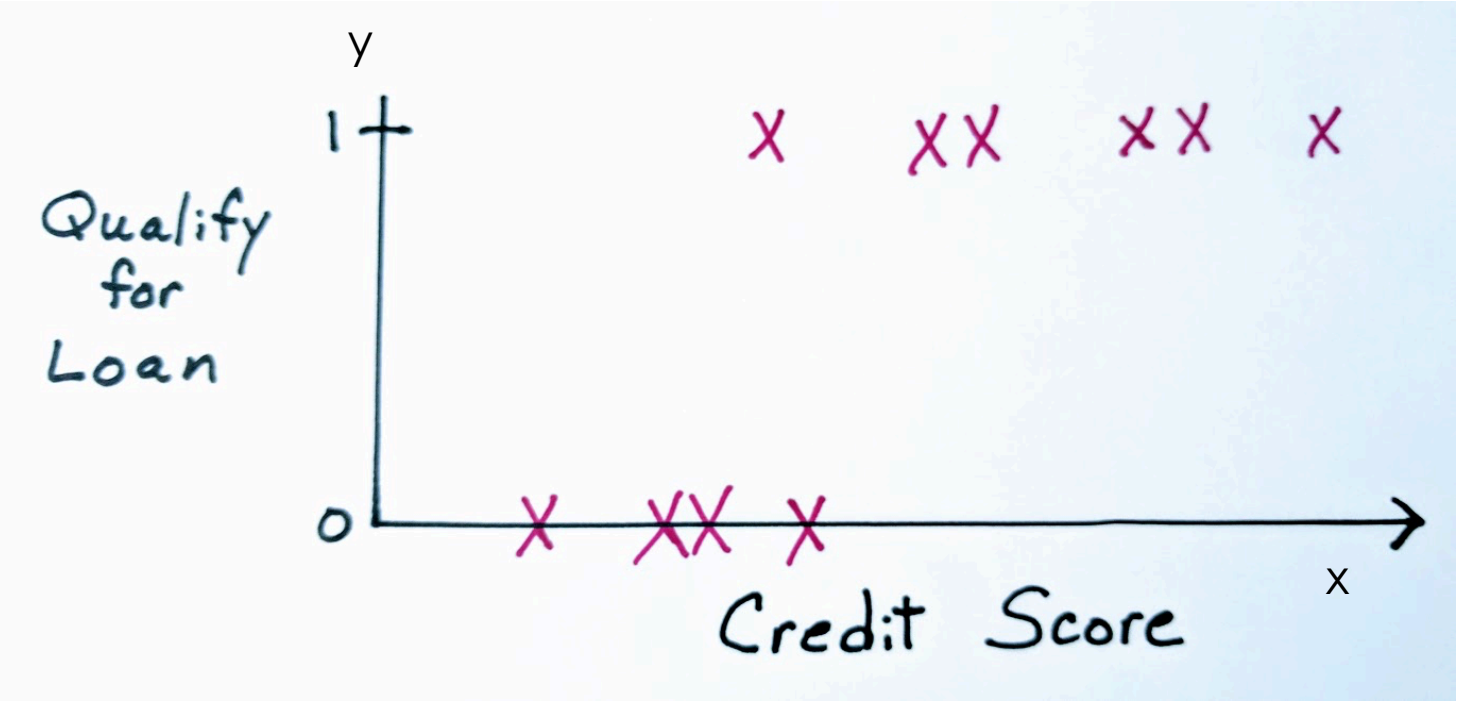


Machine Learning Algorithms

- Supervised Learning
  - Linear Regression
  - **Logistic Regression**
  - Neural Networks
- Unsupervised Learning
  - Clustering
  - Anomaly Detection
  - Recommender Systems

Logistic Regression

- Binary Classification Problem
- Model probability of being in particular class
- P(y=1 | x)

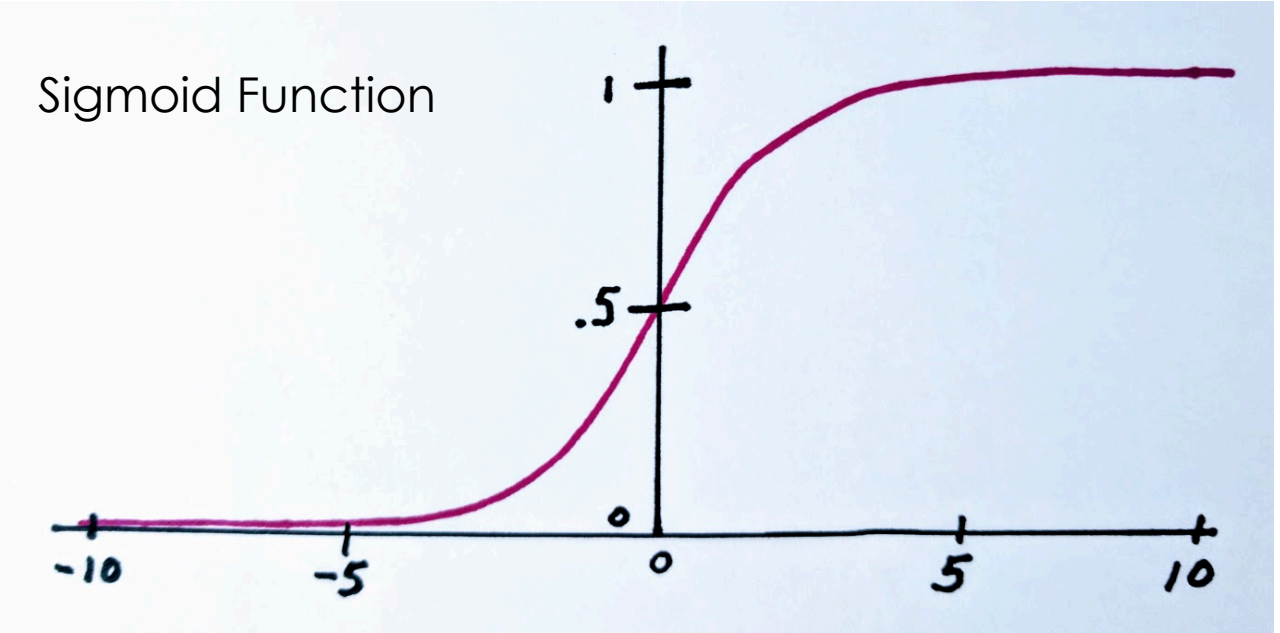


Logistic Regression

$$h_\theta(x) = g(\theta^T x)$$

$$z = \theta^T x$$

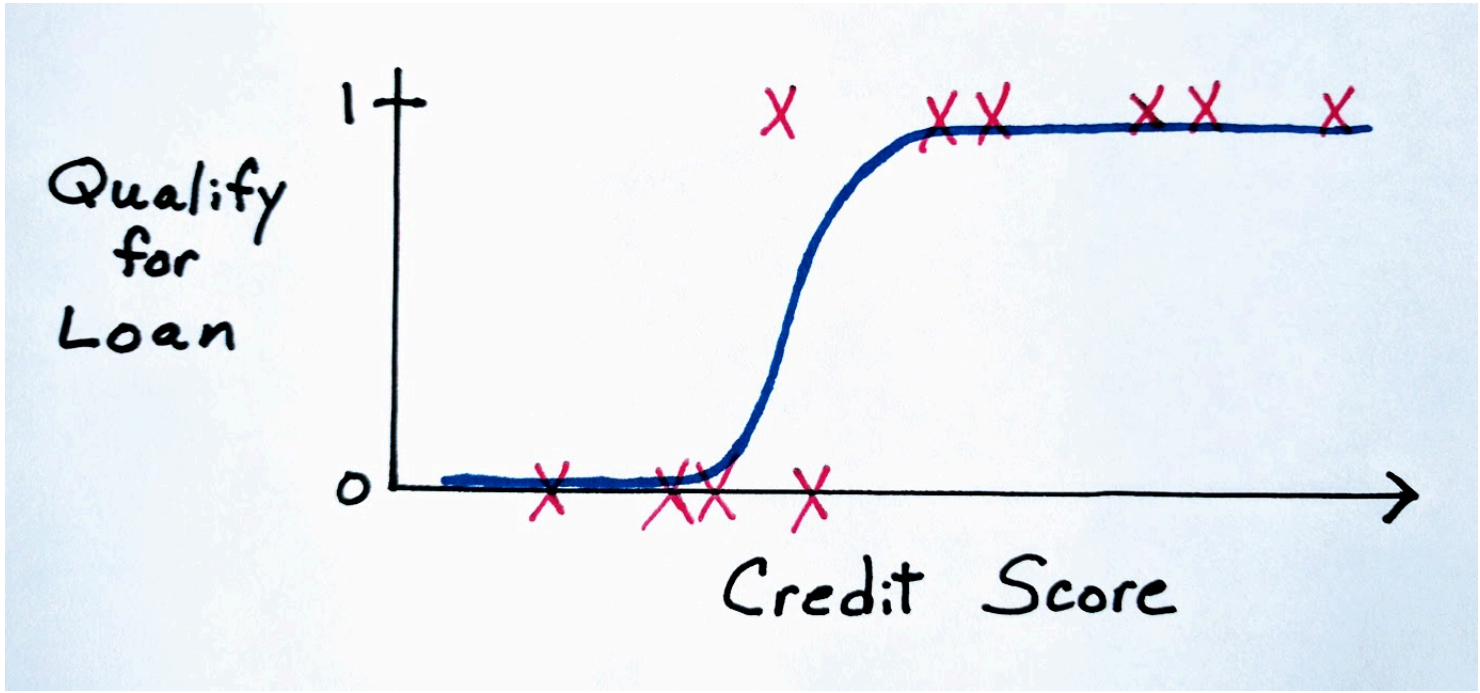
$$g(z) = \frac{1}{1 + e^{-z}}$$



$$\theta^T x = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_n x_n$$

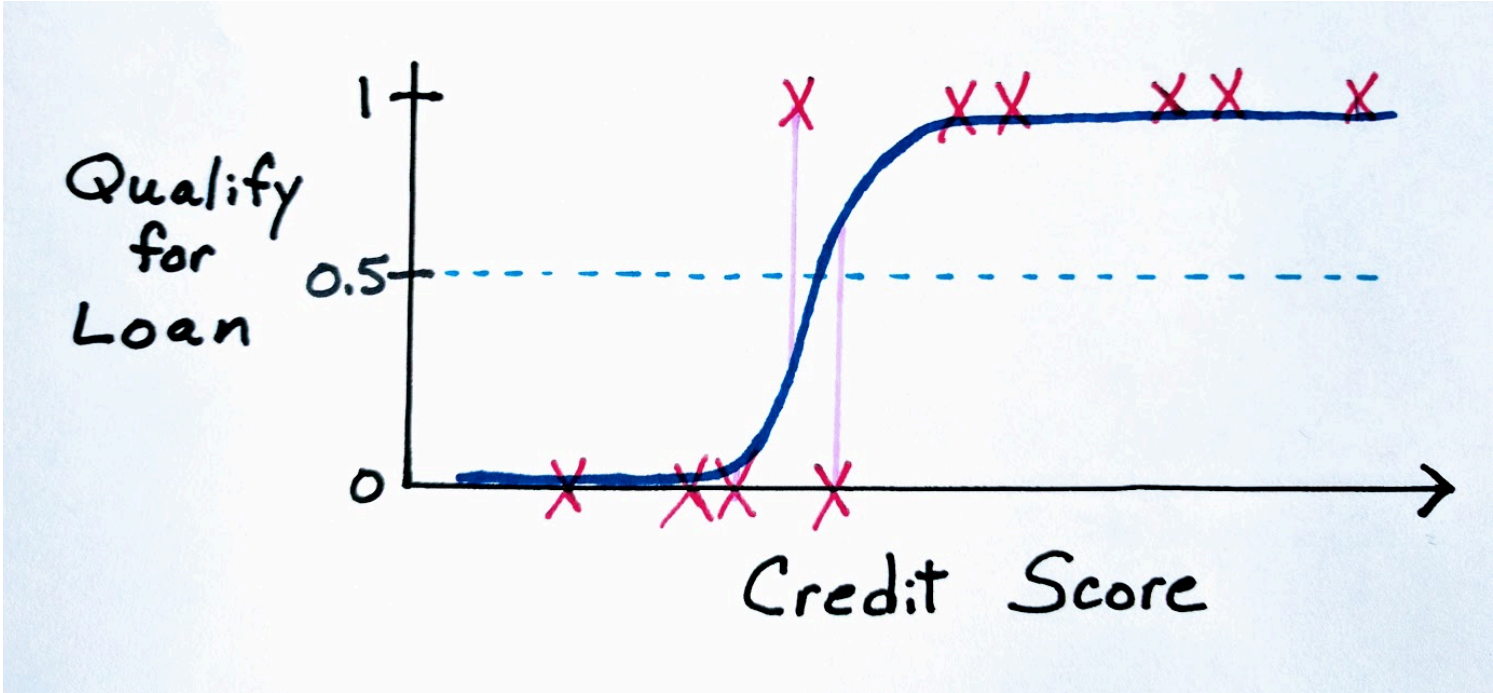
Logistic Regression

- Find best fit of sigmoid function to test data
- Estimate coefficients using Maximum Likelihood Estimate



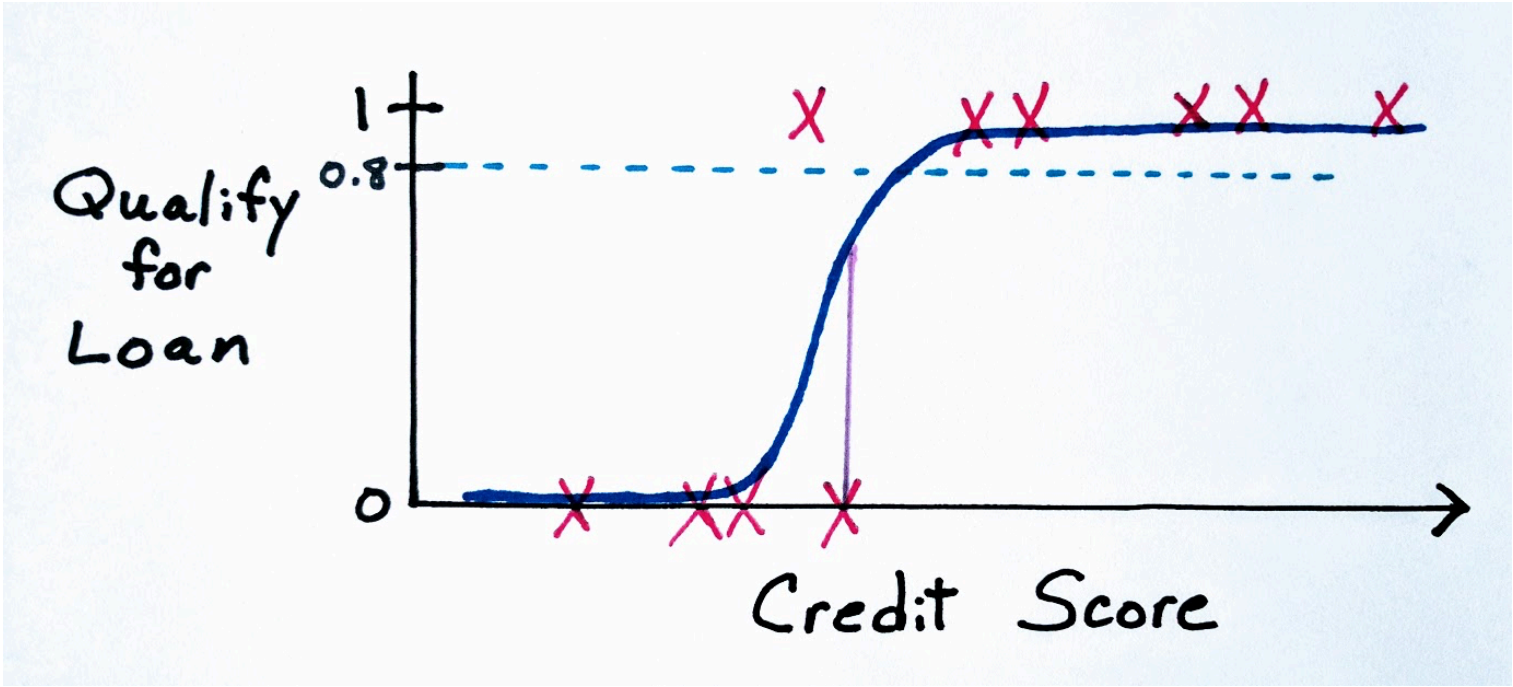
Logistic Regression

- If probability >= 0.5, output = 1 (true)
- If probability < 0.5, output = 0 (false)



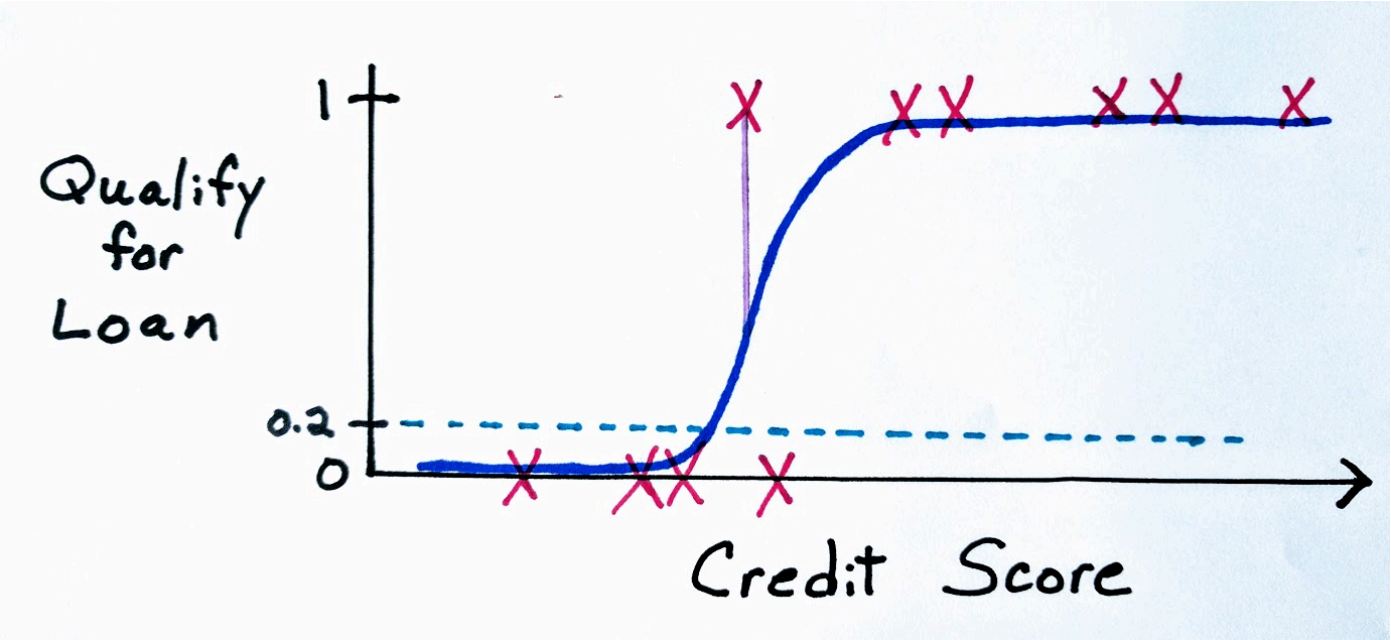
Logistic Regression

- Decision Boundary
- Use higher value (> 0.5) for higher probability of being true



Logistic Regression

- Decision Boundary
- Use smaller value (< 0.5) to include more values as true

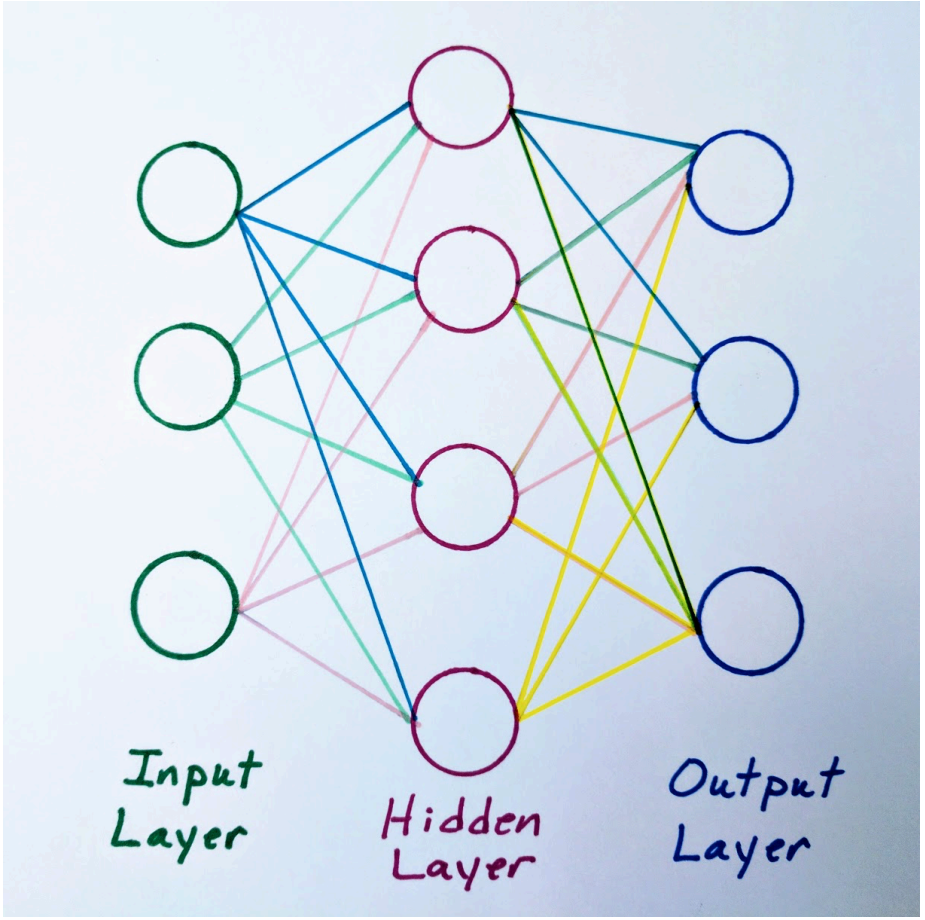


Machine Learning Algorithms

- Supervised Learning
  - Linear Regression
  - Logistic Regression
  - **Neural Networks**
- Unsupervised Learning
  - Clustering
  - Anomaly Detection
  - Recommender Systems

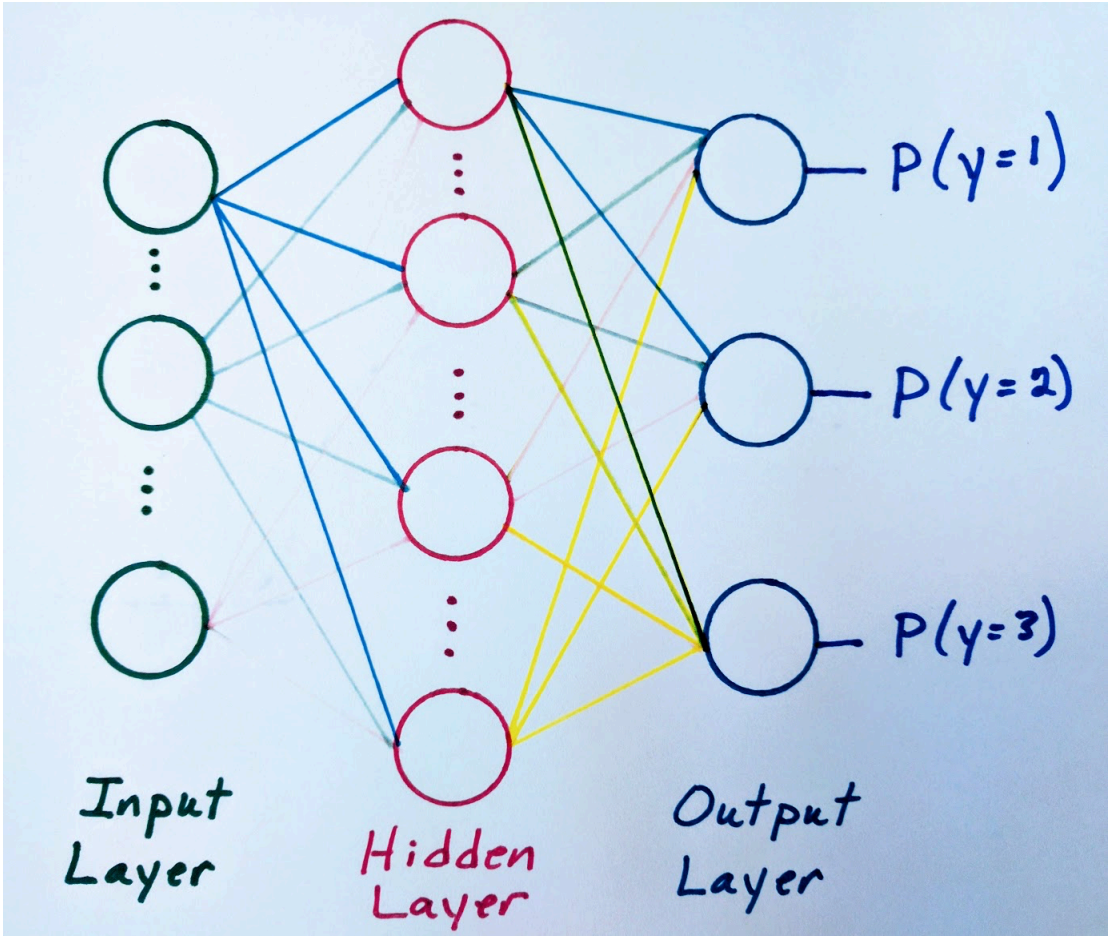
Neural Networks

- Can learn complex relationships
- Modeled on neurons in the brain
- Each node modeled as a sigmoid function or a Rectified Linear Unit (ReLU), etc.



Neural Networks

- Example: Identifying the digits 1, 2, and 3
- Input layer: pixel values of image of a digit
- Output layer: probability of digit being 1, 2, and 3



Neural Networks

- Neural network trained similar to logistic regression
- Includes a back propagation step where the error is distributed to the previous layer

Machine Learning Algorithms

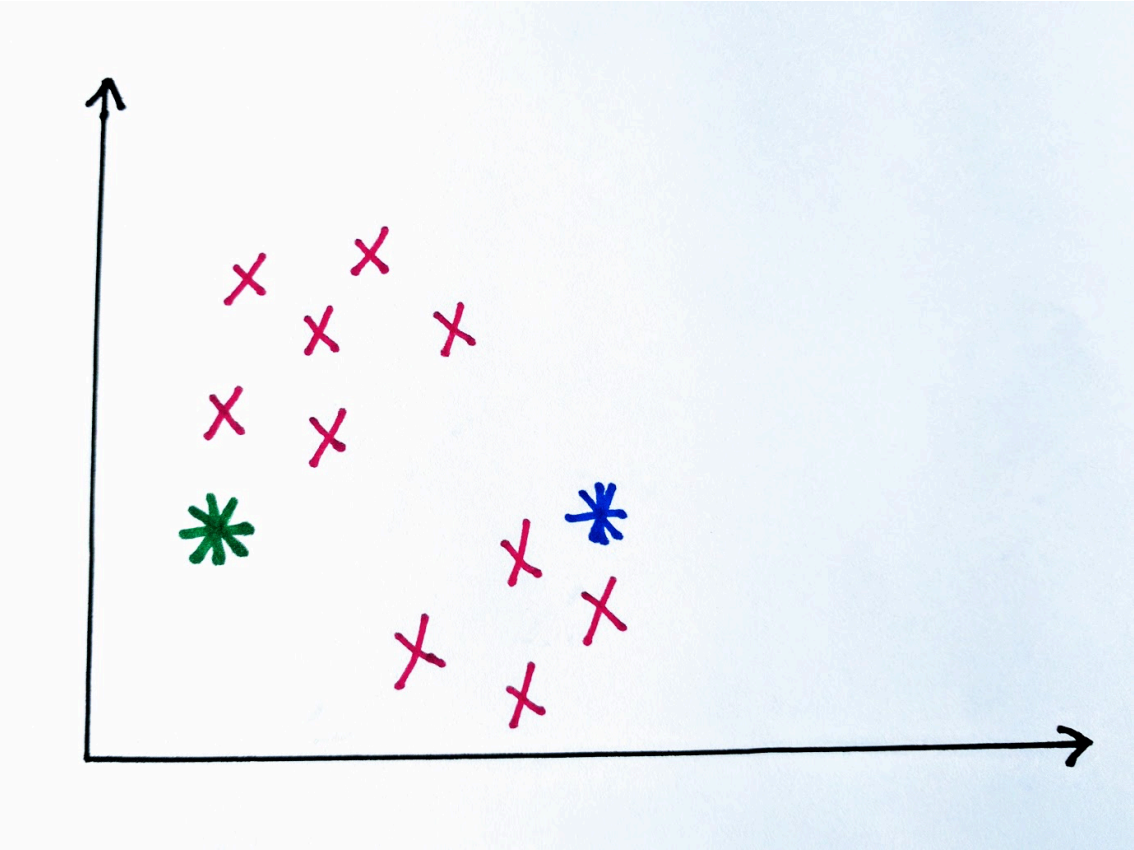
- Supervised Learning
  - Linear Regression
  - Logistic Regression
  - Neural Networks
- Unsupervised Learning
  - **Clustering**
  - Anomaly Detection
  - Recommender Systems

Clustering

- Unsupervised – data is unlabeled
- K-Means Algorithm
  - K = number of clusters/groups
  - Mean = average

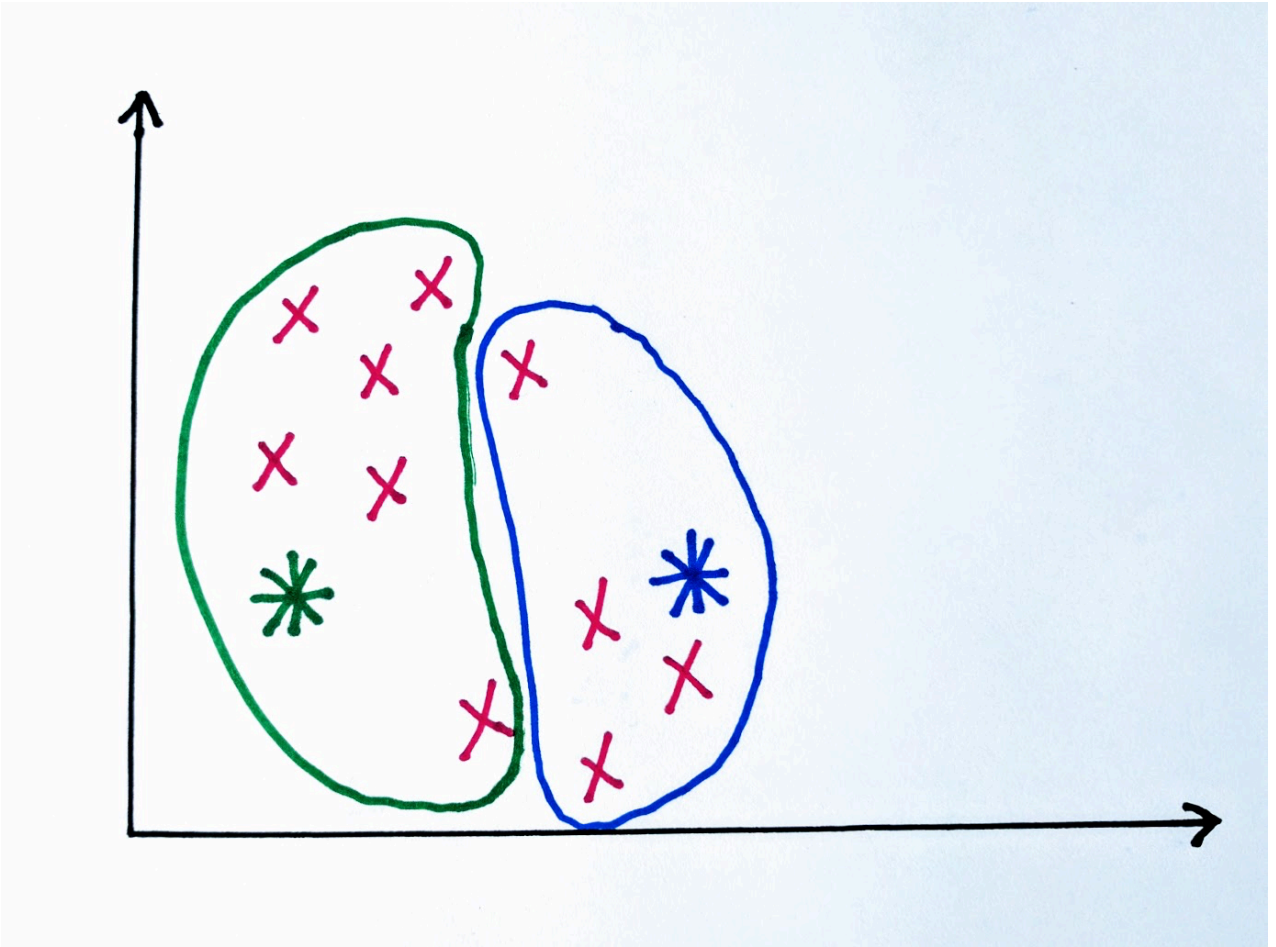
Clustering, K-Means Algorithm

- Randomly pick K points
- Use for the centers of clusters



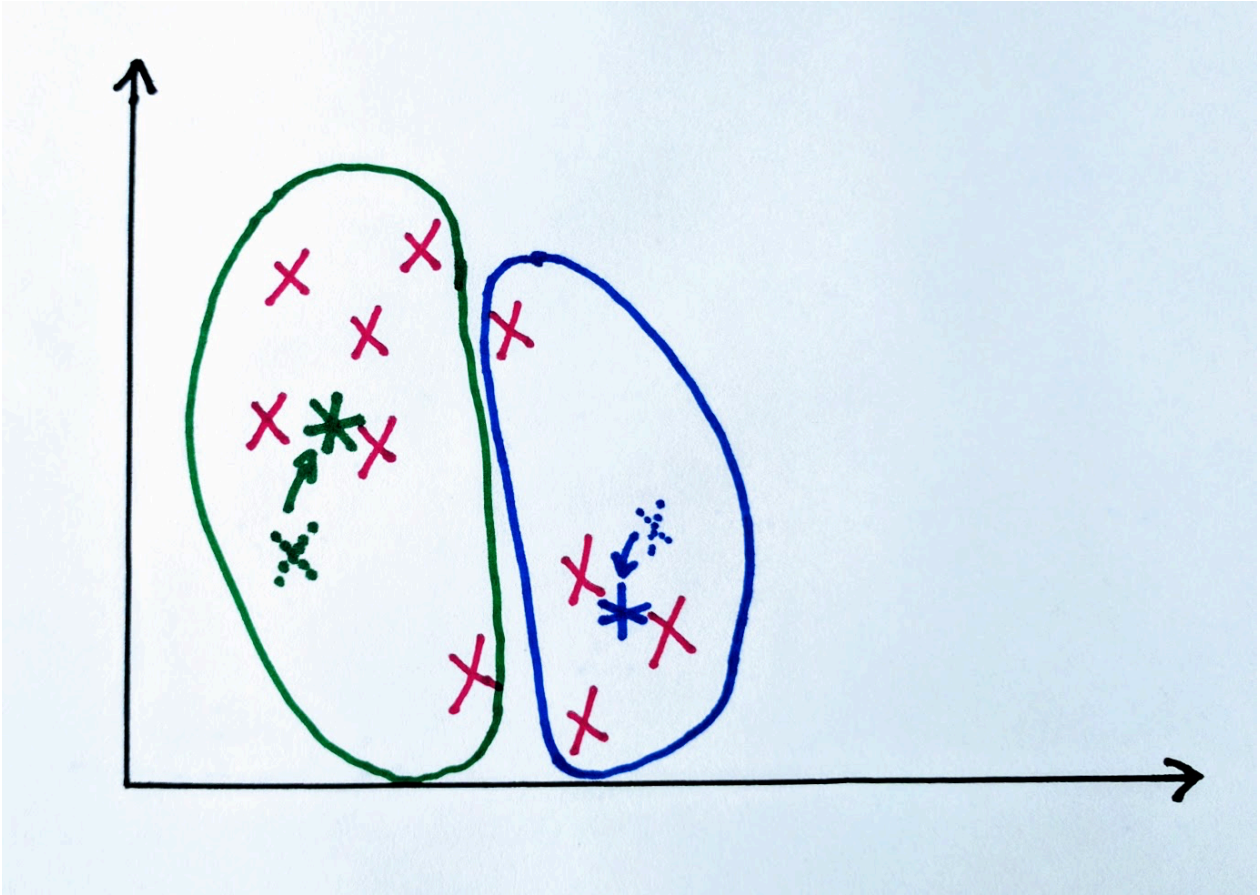
Clustering, K-Means Algorithm

- Assign each point to closest cluster center



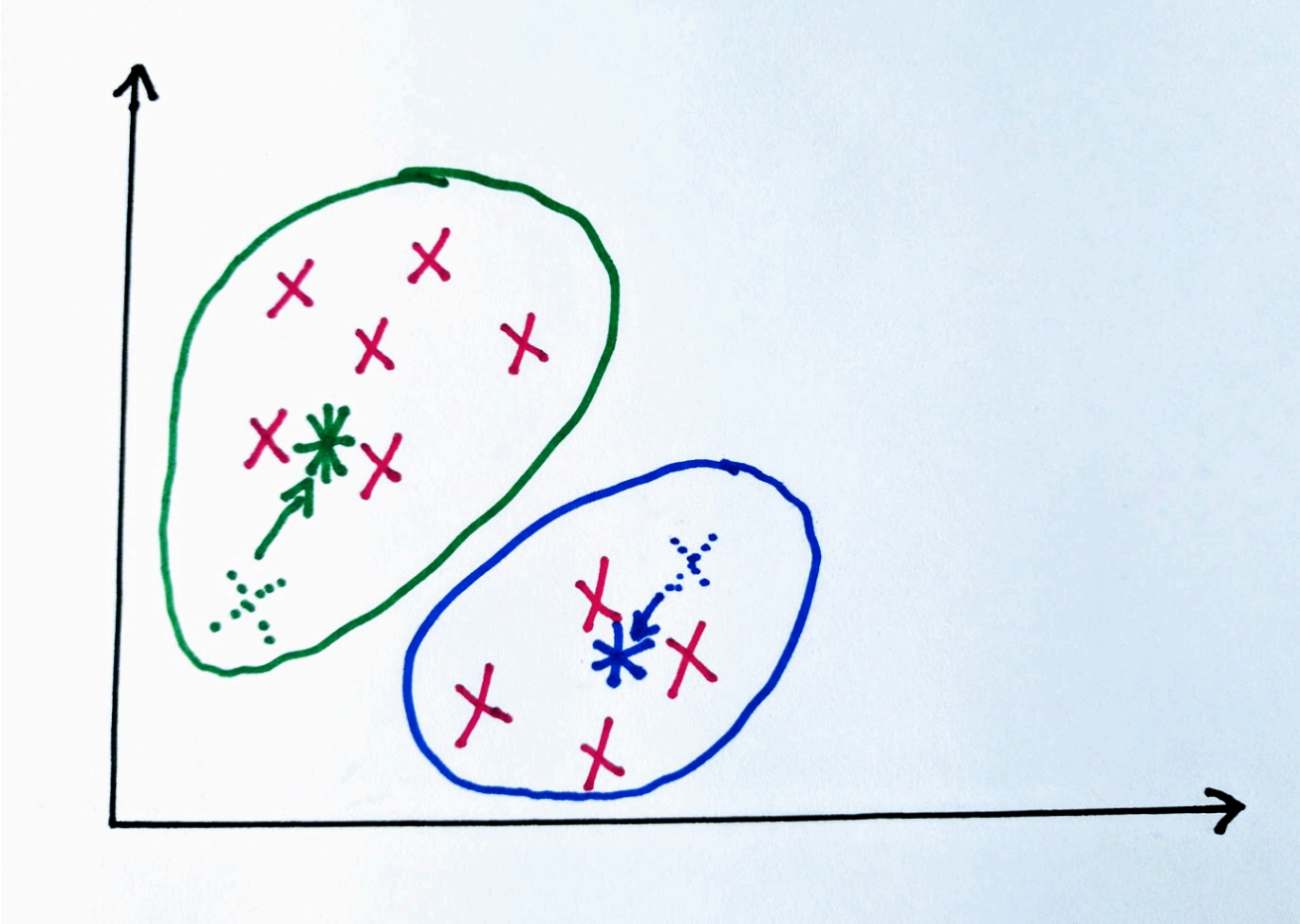
Clustering, K-Means Algorithm

- Calculate mean (average) for all points in each cluster
- Move cluster center to new mean



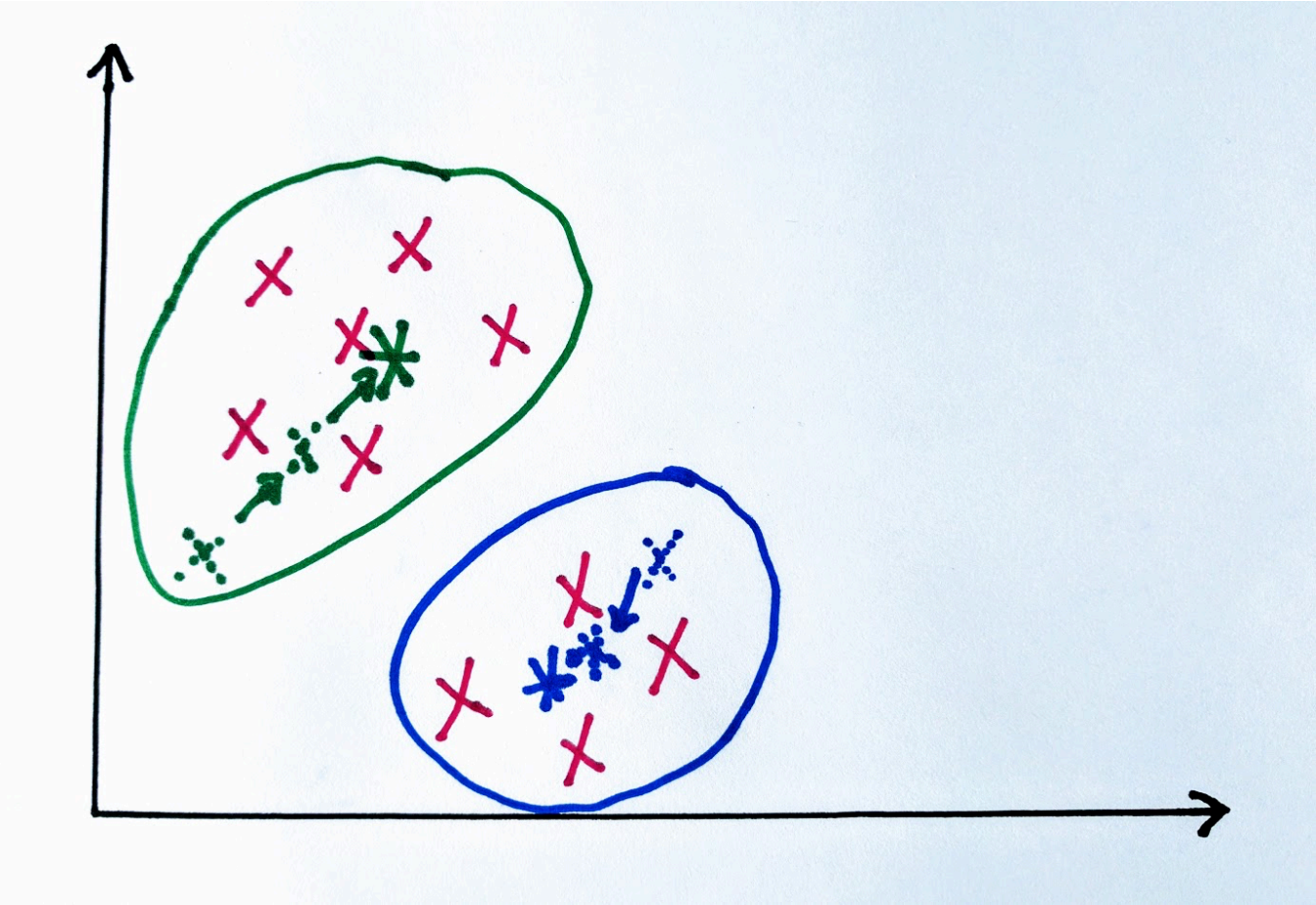
Clustering, K-Means Algorithm

- Repeat the process:
- Assign each point to new closest cluster center



Clustering, K-Means Algorithm

- Continue to repeat:
- Calculate new means
- Use new means as cluster centers
- Assign points to closest center



Clustering, K-Means

- May not converge to optimal clustering
- Try multiple random starting points
- Use actual data points for random starting points

Machine Learning Algorithms

- Supervised Learning
  - Linear Regression
  - Logistic Regression
  - Neural Networks
- Unsupervised Learning
  - Clustering
  - **Anomaly Detection**
  - Recommender Systems

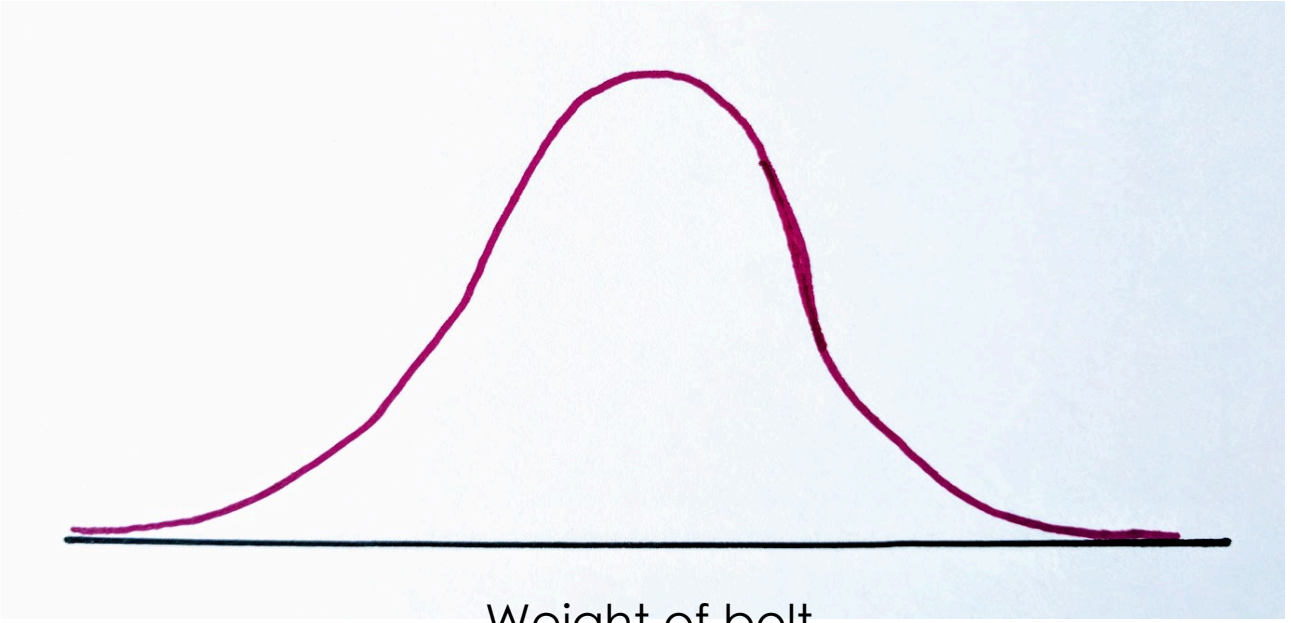
Anomaly Detection

- Detect abnormal/anomalous items
  - Defective products from assembly line
  - Attacker trying to access computer system
  - A failing computer within a data center
- If have many examples of failure, can use logistic regression
- If failure can take many forms and/or only have a few examples of failure, can use Anomaly Detection

Anomaly Detection

- Uses Gaussian Distribution, i.e. a normal distribution

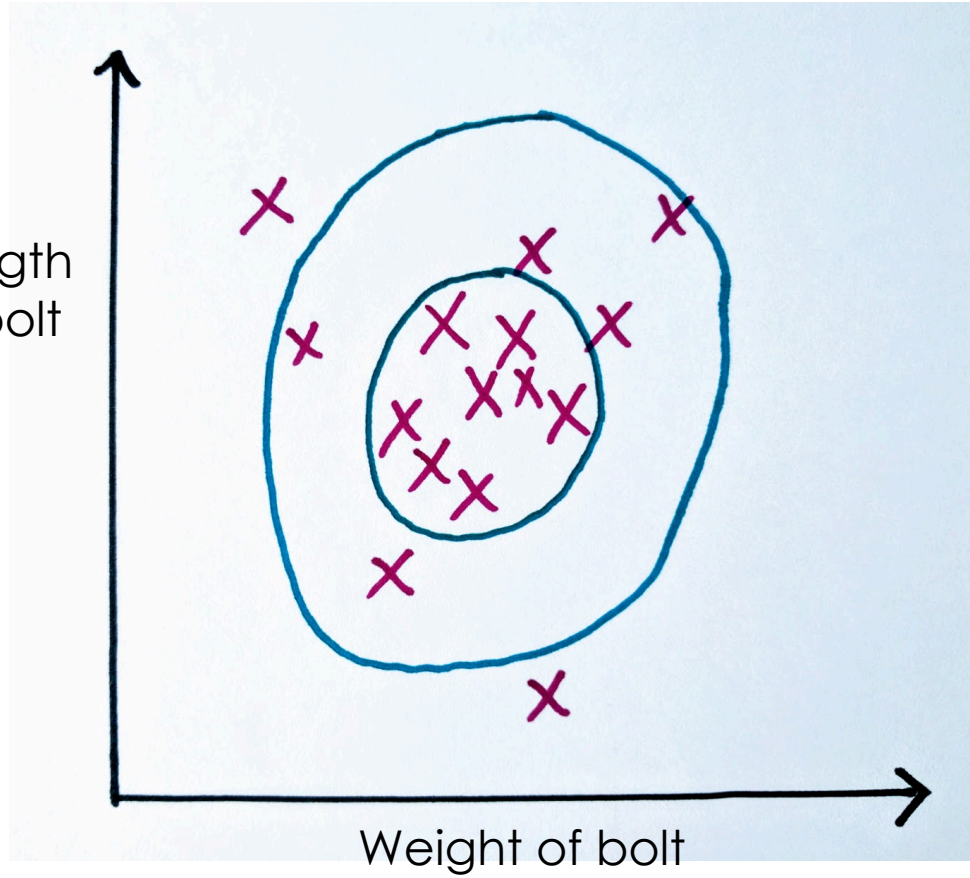
$$\frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



Weight of bolt

Anomaly Detection

- Calculate mean, mu
- Calculate standard deviation, sigma
- Determine epsilon (threshold between normal and anomalous data) with training data



Machine Learning Algorithms

- Supervised Learning
  - Linear Regression
  - Logistic Regression
  - Neural Networks
- Unsupervised Learning
  - Clustering
  - Anomaly Detection
  - **Recommender Systems**

Recommender Systems

- Content-Based Filtering
  - Characterize all products by all features
  - Example with one feature:
    - Movie A has a high value for the feature “action movie”
    - User B likes action movies
    - Recommend Movie A to User B
  - Uses linear regression to find the movies with features are the closest match to the user’s stated preferences

Recommender Systems

- Collaborative Filtering
  - Does not analyze the content of the product (such as a movie)
  - Learns the features of all products by using the rankings from all users along with those users' preferences
  - Recommend a product with highest value of feature weights * user preferences

Machine Learning

- Thank you