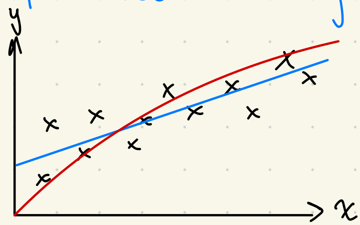
MACHINE LEARNING

Overview

ML: learn w/o being explicitly programmed

Supervised Learning



given a data set (x, y)
Map $x \to y$

Note: x is often a vector
(Multi-dimensional $\vec{x}$)

ML Strategy / Theory

– Tools to make progress more efficient/effective

Unsupervised Learning

- Data has no labels, patterns only
- Clustering algorithms
  - genetic data
  - News articles
  - Social network analysis

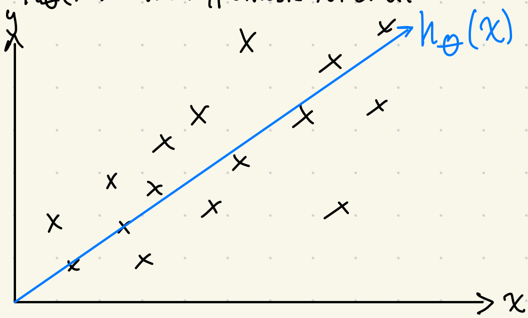
Reinforcement Learning

- There is no optimal way to do something
  - Reward for doing well
  - Scold for doing poorly

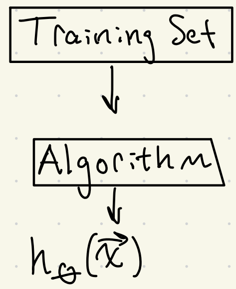
Linear Regression

Notation:

- θ = "Parameters of the learning algorithm
- m = # of training examples (# of rows)
- x = "inputs" (features)
- y = "output" (target value)
- $(\vec{x}, y)$ = a training example
- $(\vec{x}^{(i)}, y^{(i)})$ = i^{th} training example
- n = # of features
- $h_\theta(\vec{x})$ = the hypothesis function



Goal:



Normal Equations

$$\vec{\theta} = (X^T X)^{-1} X^T \vec{y}$$

$$h_{\vec{\theta}}(x) = \theta_1 + \theta_2 x$$

Stochastic Gradient Descent

$$\theta_1 \mathrel{+}= \alpha(-y_i - h_\theta(x_i))/m$$

$$\theta_2 \mathrel{+}= \alpha \cdot -x_i(y_i - h_\theta(x_i))/m$$

$$h_\theta(x) = \theta_1 + \theta_2 x$$

Locally Weighted Regression

Recall the hypothesis function:

$$h_\theta(x) = \theta_1 + \theta_2 x$$

But, we don't know if a linear function best fits the data!

New expression:

fit θ to minimize $\sum_{i=1}^{M} \exp\left(\frac{-(x_i - x)^2}{2\tau^2}\right)(y^i - \theta^T x_i)$

Logistic Regression

Used when $y \in [0,1]$

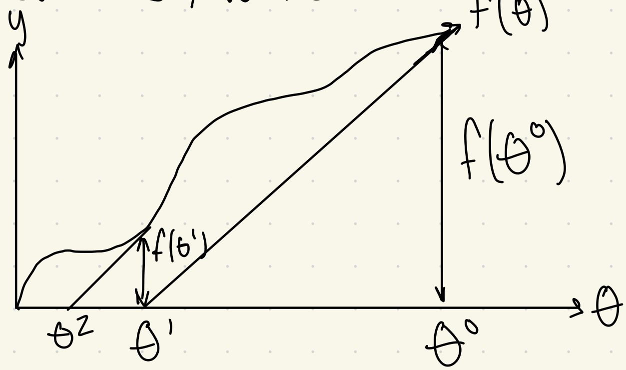
Chose θ to maximize

$$\ell(\theta) = \log(\mathcal{L}(\theta)) = \sum_{i=1}^{M} y^i \log(h_\theta(x_i)) + (1-y_i)(\log(1-h_\theta(x_i)))$$

Batch Regression

$$\theta_j := \theta_j + \alpha \cdot \sum_{i=1}^{M} y^i - h_\theta(x_i)\, x_j^i$$

Newtons Method



Update function:

$$\theta_{t+1} := \theta_t - \frac{l'(\theta_t)}{l''(\theta_t)}$$

Generitive Learning Algorithms (binary data)

Group data Seperately rather than finding a seperation boundary

Discriminative:

learn $p(y|x)$

(or $h_\theta(x) = \begin{cases} 0, & \text{blah} \\ 1, & \text{blah} \end{cases}$)

GLA:

learns $p(x|y)$

Bayes Rule:

$$p(y=1|x) = \frac{p(x|y=1)\,p(y=1)}{p(x|y=1)\,p(y=1) + p(x|y=0)\,p(y=0)}$$

Gaussian Discriminant Analysis

Suppose $x \in \mathbb{R}^n$

Assume $p(x|y)$ is Gaussian