Machine Learning Notes

(c) yungcxn

1 Introduction

ML: Tries to automate the process of **inductive inference**.

1. Deduction: Learning from rules
2. Induction: Learning from examples

2 Supervised learning

- input X, output Y
- training data: $(x^{(i)}, y^{(i)})_{i=1..n} \subset X \times Y$
- Goal: learn $f : X \to Y$ for model class F on examples

2.1 Least squares regression

$\tilde{X}, \tilde{w}$ are extended with bias:

$$\min_{\tilde{w}} \frac{1}{2} \left\| \tilde{X}\tilde{w} - y \right\|^2 \Rightarrow \min_{w} \frac{1}{2} \|Xw - y\|^2$$

Solve with gradient and set to zero:

$$\nabla L = X^T(Xw - y) = 0 \Rightarrow (X^T X)w = X^T y \Rightarrow w = (X^T X)^{-1} X^T y$$

2.2 Gradient descent

Alternative to least squares regression. Algorithm:

1. Compute gradient $\nabla L(w) = X^T(Xw - y)$
2. Negative gradient shows to steepest descent
3. $w^{(t+1)} = w^{(t)} - \gamma^{(t)} \cdot \nabla L(w^{(t)})$

2.3 Derivative examples

- $L(w) = w_1^2 + w_2^2$
  $\Rightarrow \nabla L(w) = \begin{pmatrix} 2w_1 \\ 2w_2 \end{pmatrix}$
- $L(w) = \|w\|_2^2 = w^T w$
  $\Rightarrow \nabla L(w) = 2w$
- $L(w) = w^T A w$
  $\Rightarrow \nabla L(w) = Aw + A^T w$
- $L(w) = \|Xw - y\|^2 = w^T X^T X w - y^T X w - w^T X^T y + y^T y$
  $\Rightarrow \nabla L(w) = 2X^T(Xw - y)$