Machine Learning of Andrew Ng
pdf by CarolusRex

Large scale machine learning

- Stochastic gradient descent
  - Stochastic gradient descent convergence
- Online Learning
- Map reduce and data parallelism

Stochastic gradient descent

1. Randomly shuffle(reorder) training examples;
2. Repeat { //1 - 10
   for i = 1, ..., m{
   $\theta_j = \theta_j - \alpha(h_\theta(x^{(i)}) - y^{(i)}) \cdot x_j^{(i)}$
   }
   }

Batch gradient descent: Use all m examples in each iteration;
Stochastic gradient descent: Use 1 example in each iteration;
Mini-batch gradient descent: Use b examples in each iteration.

Stochastic gradient descent convergence

Learning rate α is typically held constant. Can slowly decrease α over time if we want θ to converge. (E.g. $\alpha = \frac{const_1}{iterationNumber + const_2}$)

Online Learning

Map reduce and data parallelism