10-601 Introduction to Machine Learning

Machine Learning Department
School of Computer Science
Carnegie Mellon University

Linear Regression

Linear Regression Readings:
Murphy 7.1 – 7.3
Bishop 3.1
HTF 3.1 – 3.4
Mitchell 4.1-4.3

Logistic Regression Readings:
Murphy 8.1-8.3, 8.6
Bishop 4.3.2, 4.3.4
HTF 4.1, 4.4
Mitchell –

"Generative … Logistic Regression"
(Mitchell, 2016)
"Maximum … Gradient Training"
(Elkan, 2014)

Matt Gormley
Lecture 8
February 13, 2016

Reminders

- Homework 2: Naive Bayes
  - Release: Wed, Feb. 1
  - Due: Mon, Feb. 13 at 5:30pm
- Homework 3: Linear / Logistic Regression
  - Release: Mon, Feb. 13
  - Due: Wed, Feb. 22 at 5:30pm

Linear Regression Outline

- **Regression Problems**
  - Definition
  - Linear functions
  - Residuals
  - Notation trick: fold in the intercept
- **Linear Regression as Function Approximation**
  - Objective function: Mean squared error
  - Hypothesis space: Linear Functions
- **Optimization for Linear Regression**
  - Normal Equations (Closed-form solution)
    - Computational complexity
    - Stability
  - SGD for Linear Regression
    - Partial derivatives
    - Update rule
  - Gradient Descent for Linear Regression
- **Probabilistic Interpretation of Linear Regression**
  - Generative vs. Discriminative
  - Conditional Likelihood
  - Background: Gaussian Distribution
  - Case #1: 1D Linear Regression
  - Case #2: Multiple Linear Regression

Last Lecture

This Lecture

Regression Problems

Whiteboard

- Definition
- Linear functions
- Residuals
- Notation trick: fold in the intercept

Linear Regression as Function Approximation

Whiteboard

- Objective function: Mean squared error
- Hypothesis space: Linear Functions

Optimization for Linear Regression

Whiteboard

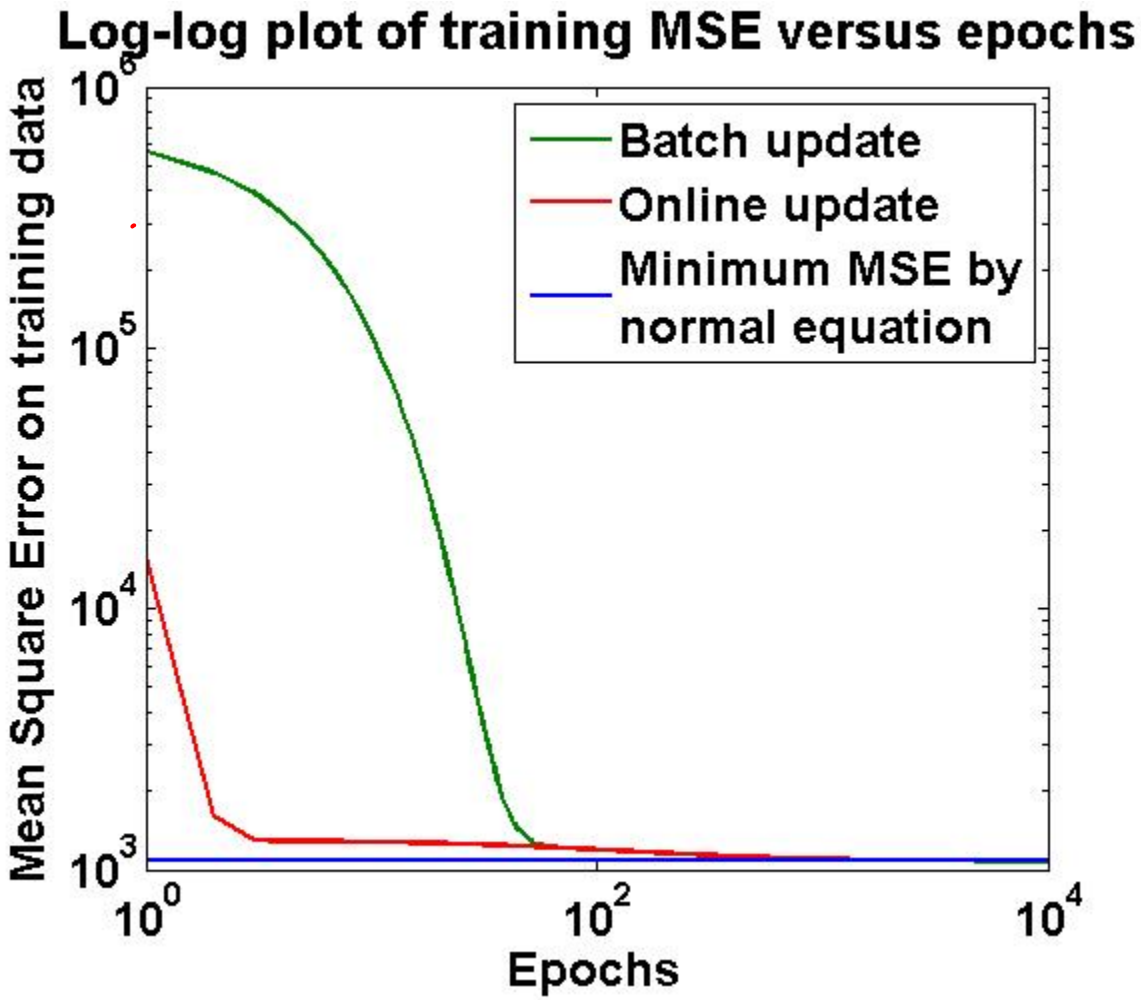
- Normal Equations (Closed-form solution)
  - Computational complexity
  - Stability
- SGD for Linear Regression
  - Partial derivatives
  - Update rule
- Gradient Descent for Linear Regression

Probabilistic Interpretation of Linear Regression

Whiteboard

- Generative vs. Discriminative
- Conditional Likelihood
- Background: Gaussian Distribution
- Case #1: 1D Linear Regression
- Case #2: Multiple Linear Regression

Convergence Curves



- For the batch method, the training MSE is initially large due to uninformed initialization
- In the online update, N updates for every epoch reduces MSE to a much smaller value.