Lesson 07
Notes on Chapter 4
Gradient Descent

CSC357 Advanced Topics—Machine Learning

24 January 2020

- Gradient Descent is very different way to train a Linear Regression model
- better suited for cases where there are...
  - a large number of features
  - or too many training instances to fit in memory
- a generic optimization algorithm
- capable of finding optimal solutions to wide range of problems
- tweak parameters iteratively in order to minimize a cost function
- suppose you are lost in mountains in dense fog
- can only feel the slope of the ground below your feet
- good strategy: get to bottom of valley quickly by going downhill in direction of steepest slope
- measures local gradient of error function with respect to parameter vector θ
- goes in direction of descending gradient
- once the gradient is zero, you have reached a minimum!
- start by filling θ with random values (*random initialization*)
- then, small steps
- at each step, try to decrease cost (MSE)
- learning step size proportional to slope of cost function