s

**Lesson 04**

**CSC357 Machine Learning**

**21 January 2020**

**Stochastic Gradient Descent**

**Instructor:** **Leon Tabak**

**Authors:** **Alex Heisdorffer, Chenxi Liu, Yudong Zhang, Linda Wang**

Gradient Decent & Stochastic Gradient Decent

Gradient Decent:

**Gradient descent** is an optimization algorithm used to minimize some function by iteratively moving in the direction of steepest **descent** as defined by the negative of the **gradient**. Dependent variable is error.

Linear Regression

Loss Function (In the linear regression): Sum of the residuals = (Observed Height – Predicted Height) ^2

1. Take derivatives of the loss function respect to intercept and slope
2. Plug the initial guess of intercept and slope, get new slope2 at a particular point
3. Step Size = slope2 \* learning rate
4. New Intercept = Old Intercept – Step Size

New Slop = Old Slope – Step Size



Gradient Decent stops when the step size is very close to 0.

Every time, we will have better result: We can have new intercept and new slope.

Stochastic Gradient Decent:

In both gradient descent (GD) and stochastic gradient descent (SGD), you update a set of parameters in an iterative manner to minimize an error function.

What happen if we have complicated model?

Advantage:

1. SGD would randomly pick on sample for each step.
2. Just use that one sample to calculate the derivatives.
3. In this example, SGD reduced the number of terms computed by a factor of 3.

Stochastic Gradient Decent is especially useful when there are redundancies in the data.

1. SGD normally means using one sample per step.
2. It is more common to select a small subset of

data, or Mini-Batch.

1. Mini-Batch is more stable and faster.

Reference:

<https://ml-cheatsheet.readthedocs.io/en/latest/gradient_descent.html>

<https://towardsdatascience.com/stochastic-gradient-descent-clearly-explained-53d239905d31>

<https://developers.google.com/machine-learning/crash-course/reducing-loss/stochastic-gradient-descent>