



An Introduction to Machine Learning

Part 2: Linear Regression



Recall

Most Machine Learning algorithms can be described as a combination of the following:

- Dataset
- Cost/Loss Function
- Optimization Procedure
- Model

Linear Regression

- Linear Model of our data
- Example of Supervised Learning
- A gift from Statisticians

Dataset

In linear regression our data sets should be in the form of a set of feature vectors X and their correlated labels Y .

Examples:

- Cricket Chirp Rate and Temperature
- Wine Characteristics and Rating

Model

Some common Linear Regression Models are:

- $Y' = wx$
- $Y' = wx + b$
- $Y' = w * x$
- $Y' = w * x + b$

Y' is the predicted label, w is the weight scalar/vector, x is the feature scalar/vector and b is a bias value.

Cost Function

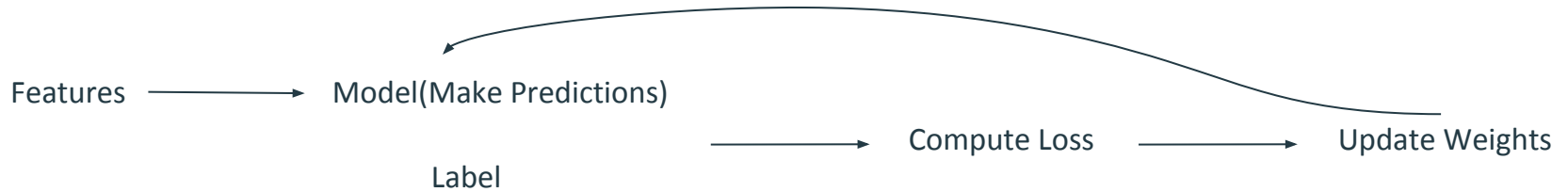
Arguably the most common cost function used in linear regression models is Mean Squared Error.

We want to use the error of our Training data to minimize the generalization error.

One approach is to simply minimize the error on the training set.

Beware of Overfitting!

Optimization Procedure



Optimization Procedure

Some of the most common methods of optimization are variants of the gradient descent algorithm.

The procedure is an iterative process.

Gradient Descent

Step 1: Set initial guess for our weights (usually 0 or a random value.)

Step 2: Calculate the gradient of the loss curve for our initial guess.

Step 3: Subtract a scalar multiple of the gradient from our initial guess.

Step 4: Repeat process as needed.

Gradient Descent

- Local Minimums
- Vanishing Gradients
- Data Size
- Learning Rate

Variations on Gradient Descent

Stochastic Gradient Descent (SGD):

- Uses only one example chosen at random per iteration.
- Works given enough iterations, but is very noisy.

Mini-batch stochastic gradient descent (mini-batch SGD):

- A group of examples is chosen at random during each iteration.
- Size is typically between 10 and 1,000 examples.
- Less noisy than SGD but more efficient than full-batch.

Resources

<http://www.deeplearningbook.org/>

<https://developers.google.com/machine-learning/crash-course/>

<https://www.python.org/>

<https://www.hackerrank.com/>

https://www.amazon.com/Elements-Statistical-Learning-Prediction-Statistics/dp/0387848576/ref=sr_1_4?ie=UTF8&qid=1539691739&sr=8-4&keywords=statistical+learning+theory