Google Machine Learning Crash Course Cheat Sheet

by furkandurul via cheatography.com/122404/cs/22715/

Machine Learning Terminology

Label is variable we're predicting. Represented by **y**.

Features are input variables describing data. Represented by the variables **{x1,x2,...,xn }**

Example is a particular instance of data, **x**
- **Labeled example** has {features, label}: (x,y)
  Used to train the model.
- **Unlabeled example** has {features,?}: (x, ?)
  Used for making predictions on new data.

Model maps examples to predicted labels: y'. Defined by internal parameters, which are learned.

Training means creating or learning the model. You show the model labeled examples and enable the model to learn the relationship between features and label.

Inference means applying the trained model to unlabeled examples. You use the trained model to make useful predictions (y').

Regression model predicts *continuous values*.
For example; What is the value of a house in California?

Classification model predicts *discrete values*.
For example; Is a given e mail message spam or not spam?

Hyperparameters are the knobs that programmers tweak in machine learning algorithms.

Model and Equation

Equation for a model in machine learning;
  y'=b+w1x1

y' is the predicted label.
b is the bias, also referred to as w0.
w1 is the weight.
x1 is a feature (a known input).

Some models have multiple features. For example, a model relies on three features look as follows;
  y'=b+w1x1+w2x2+w3x3

Training and Loss

Training a model means learning values for all the weights and bias from labeled examples.

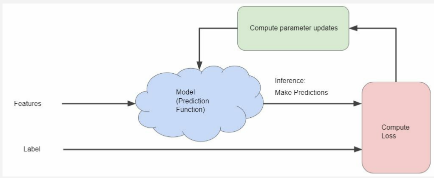
Loss is a number indicating how bad the model's prediction on a single example. The goal of training a model is to find a set of weights and biases that have low loss, on average, across all examples.

Training and Loss (cont)

Mean square error (MSE) is the average squared loss per example over the whole dataset.

MSE=1/N ∑ (x,y) ∈D (y-prediction(x))^2
- x is set of features.
- y is example's label.
- prediction(x) is function of the weights and bias of features of x.
- D is data set containing labeled examples.
- N is the number of examples in D.

Reducing Loss



Reducing Loss

Learning continues iterating until the algorithm discovers the model parameters with the lowest possible loss. Usually, until overall loss stops changing or at least changes extremely slowly. When that happens, we say that the model has **converged**.

Gradient descent algorithm calculates the gradient of the loss curve. When there is *single weight*, gradient of the loss is the derivative (slope) of the curve, When there are *multiple weights*, the gradient is a vector of partial derivatives with respect to the weights.

Gradient is a vector, so it has both of the following characteristics; a *direction* and a *magnitude*
The gradient always points in the direction of steepest increase. The gradient descent algorithm takes a step in the direction of the **negative gradient** in order to reduce loss as quickly as possible.

Gradient Descent