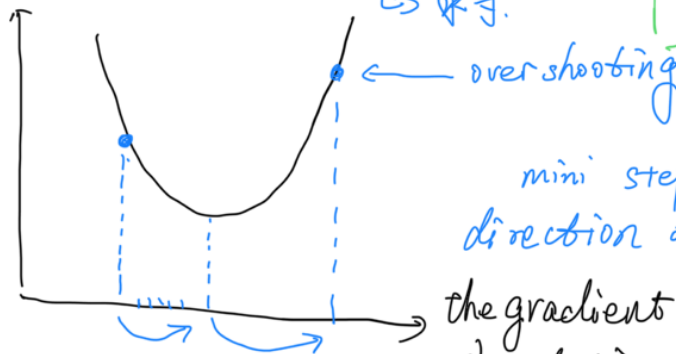


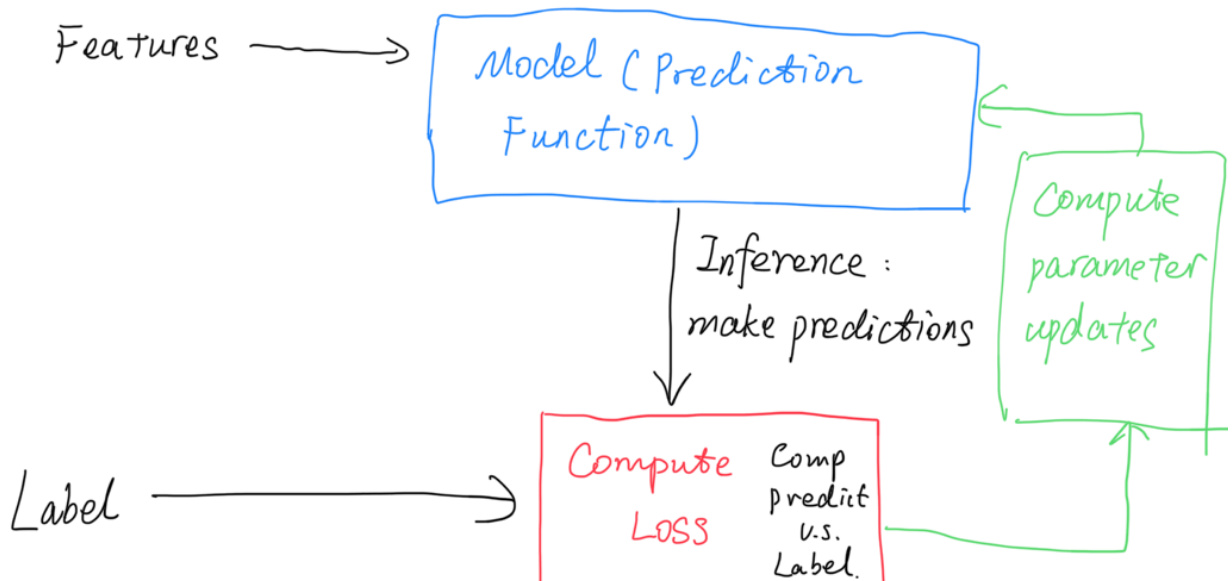
reducing loss \Rightarrow gradient 梯度
 \hookrightarrow 求导.

gradient descent
梯度递减



mini steps to find which is the
direction of minimizing loss
the gradient of the loss is equal to
the derivative (slope) of the curve

Trial-and-error process that machine learning algorithms
use to train a model:



gradient is a vector. so it has both of the following characteristics: ① direction ② magnitude ^{multiplier}

Learning rate / Step-size \Rightarrow this is usually a percent.
small learning rates take forever. Large learning rates cause overshooting.

batch is the total number of examples you use to calculate the gradient in a single iteration

Choosing examples at random from data set \Rightarrow Stochastic gradient descent (SGD): extreme: only 1 random example

Minibatch SGD: 10 ~ 1000 examples per iteration

Current hierarchy of Tensorflow:

Tensorflow Estimators \leftarrow high level . oob API

tf.layers, tf.losses, tf.metrics \leftarrow reusable libraries for common model components

Python tensorflow \leftarrow provide ops, which wrap C++ kernels

C++ TensorFlow

CPU GPU TPU \leftarrow Kernels work on one or more platforms

TPU: TensorFlow Processing Unit

tf.estimator is compatible with scikit-learn API

Types of data we are dealing with:

Categorical data & Numerical data

data that is textual

number / value

文本处理数据

integer / float

e.g. home style.

words in real-estate ad.

feature column: store the description of the feature data, not the data itself.