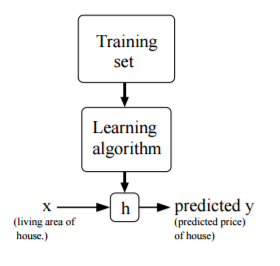
Gradient Descent:

Basically have a dataset which maps inputs to outputs. Example: Living Area vs. Price in Portland. You want to predict the relationship for values other than what you are given.

The given dataset is a training set.

|  |  |
| --- | --- |
| xi | Yi |
|  |  |
|  |  |
|  |  |



h is a hypothesis function which is what you are trying to get. It has some coefficients in it.

Linear Regression: Assume a linear function of all the parameters which might affect the predicted y. So if housing prices might depend on square footage and # rooms, then maybe

y = a + b (square footage) + c ( # rooms).

So y is a linear superposition of the independent variables plus a y-intercept.

In a linear regression, we change the coefficients a, b and c in order to minimize a “cost” function which is

J(a,b,c)= 1/2

Basically we’re evaluating our hypothesis function y for each value of the training set (summing over i) and then comparing that to the training set.

Learning Algorithms:

1. LMS or Widrow-Hoff… Coefficients are changed by some constant times the difference between the evaluated yi and the training Yi. Just like a kp in PID.