Machine learning lecture 1

Linear regression function:

R=a0+a1\*x1+...+an\*xn, where n means the number of features.

vectored notation:

A = [a0, a1, ... an]

X = [x0, x1, ... xn], where x0 always = 0;

So R = A \* X

# Supervised learning

# Regression problem

Linear regression

R = A \* X, for one feature, R = a0 + a1 \* x1;

The goal is to find best a0 and a1 to best fit x1 (minimize the cost function = minimize the sum of absolute value or power, (because of the negative values) between the estimated data and correct data)

# Cost function:

Linear function:

Where n is the number of features, h(xi) is the evaluated date and yi is the correct data.

Gradient descent (GD)

Repeat until regression

Update:

is the start point, is the learning rate( is important, too big is easy to miss the minimized value while too small equal to too many steps).

So the goal is to find the minimized value

So

\*\*\*\* =

# Go batch GD

(1) iteration

(2) temp

temp

temp

this is only one iteration

# Stochastic GD

repeat until

# Stochastic GD & Batch GD

Examples:

(x1,y1), (x2,y2),…,(,)

Xi = [

So compute matrix:

X = [X1’ ; X2’ …… Xm’];

y = [

y = X \*

# Classification

Logistic regression

A small number of discrete values.

Example

Email spam or non-spam

y is the variable that use to predict:

y = {0,1} where 0 means non-spam, 1 means spam.

Class function:

对g求导

g’= g(z)(1-g(z))