GRADIENT DESCENT

1)CLASS:

Categories model o/p

o/p --Dependent var

Person identity   male or female

2) Features(attributes)

Char define our pb

Data points-----i/p

i/p----independent var

features---height,weight,foot size

3)parameters

Estimates of true distr of whatever we r trying to classify

Values ctrl the learning process

Adjusting param values ---accurate model

Also determines the val of **model parameter**.

Hyper parameter----top level parameter—ctrls learning pr+model param

Learning rate,no. of epochs,no of clusters…

Model parameters---determined uzsing the training dataset.

Are the fitted parameters.

Eg: linear regression model using m-dim training dataset

Yi=

X=predictor matrix

W=weights------w0,w1,w2……model parameters

Learning rate,no iterations-----hyper param

→ Learning rate:
→ Hyper param---how much to change the model in response to the estimated error---each time the model wts are updated…

→cost function(loss fun/error metric)

Determines the perf of the model.

Diff between the predicted val n the actual val

Several iterations…

Lesser the val of cost fn-----better the model.

Types

1) Dist based error
2) Mean sqrd error
3) Root mean sqrd error
4) Mean absolute error
5) Cross entropy fn
   Measures dist between two prob distr p n q
   p----actual prob distr
   q----predicted prob distr.

   $H(x)=-\text{sigma } i=1 \text{ to } N \, p(x) \log q(x)$

N=no of observ taken

p(x)=prob distr of actual val

q(x)=prob distr of predicted val

→Gradient Descent

Optimization tech

Gradient??? Slope of a fn---measures the <u>degree of change of a variable</u>

Slope=0 model stops learning pr

3 types

1) **Batch GD**
2) **STOCHASTIC GD**
3) **MINI GD**

1)<u>BATCH GD(VANILLA GD)</u>

Calculates the error for each eg/samples within the training dataset.

Training epoch..

2) <u>stochastic gd</u>

A few samples

Batch-----total no of samples ---for calculating the gradient for each iteration.

Batch gd-----batch=whole dataset,find out the gradient---sum of the gradient of cost fn of all the samples.

In SGD-----gadient of the cost fn of a single sample at each iteration…..

4) **<u>Mini gd</u>**

   Stochastic+batch gd

   Splits the training dataset into small batches

   Batch size-range 50 to 256..

   Go-to mthed/algo----deep learning