

Linear Regression → It is technique to find relationship between features & labels or Statistical technique to find relationship between variables

Hyperparameters

↳ There are variables which controls different aspects of training

- | | |
|-------------------|--|
| ① → Learning Rate | } There are hyperparameters are control by the user/programmer/developer/myself. |
| ② → Batch Size | |
| ③ → Epochs | |

⇒ Parameter → All the parameter in model like bias and weights which is calculated by the Model itself while training but hyperparameters are one mostly controlled by developer/user

⇒ ① Learning Rate

⇒ Learning Rate is floating pt number

⇒ Lower the Learning Rate the model will take larger time to converge

⇒ Higher the Learning Rate it bounces between bias & weights that minimising the loss. and never converges

⇒ Goal is to pick the good/better learning rate so it converges quickly.

⇒ Learning Rate determines the magnitude of change to make to the weights & bias so the gradient descent would get weights & bias appropriate so model converges quickly

⇒ It multiplies the gradient by the learning rate to determine the model parameter (bias & weight)

⇒ third step of gradient descent, the 'small amount' to move in a ~~negative~~ direction of negative slope is 'sferred'

our learning rate.

⇒ Difference between old & new model parameter is proportional to the slope of the Loss function

⇒ If slope is large → Model takes large step

⇒ If slope is small → It takes small

Batch Size → Hyperparameter which tells about, no. of examples model processes before updating its weights & bias

Two techniques to get right gradient on average without getting all the examples to process

① Stochastic gradient descent (SGD)

② Mini-batch Stochastic gradient descent.

① SGD → This technique uses one single example (i.e. batch size = 1) per iteration

→ Given enough iterations it will work with

but lot of Noise



Noise refers to variations during training data which causes the Loss to increase rather than decrease.

STOCHASTIC → One example comprising each batch chosen randomly

② Min-Batch SGD

↳ It's a compromise between SGD &

Full Batch SGD

⇒ for a no. of datapoints → Batches can be greater

Bas 1 & Ex Plan 1

- ⇒ Model chooses examples included in each batch at random and average their gradients. update the weights & bias per iteration
- ⇒ Determining the number of examples for each batch depends on dataset & the available compute resource.
- ⇒ Noise is not always used it help a model to generalize weights & bias.

EPOCHS

→ Epoch tells the unit about model has complete all the examples once.

Eg: There are 1000 example in training set and mini-batch size 100 examples. It takes 10 ^{iterations} ~~epochs~~ to complete the ~~examples~~ 1 epoch.

⇒ Training basically need many epochs

→ It means system needs multiple to process the example multiple time.

⇒ parameter (bias & weights) are updated every epoch.

same for 20-epochs → model update parameters 20 times