Hyper parameters

Friday, 25 June 2021 7:38 PM

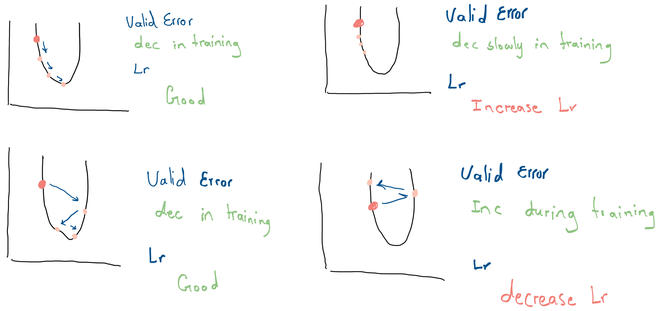
-Optimizer Hyper parameter

- Lr , Minibatch size , Epochs

-Model Hyper param. / involved in structure of model

– # layers , hidden units

★ Learning Rate :



Each W_i has its own Error Curve

Learning rate decay will dec the value of Lr as the model trains

* Linear Lr decay
* Exponential Lr decay

or use adaptive learning

* Adam optimizer
* Adagrad optimizer

} inc or dec Lr adaptivily as the model learns. (when needed)

★ Minibatch size

(Affects resource requirments + training speed)

Recomended :

Potential good

1, 2, 4, . . . , 128, 256,

too slow — 32 often good ! — too taxing computationaly

Small minibatch: has noise in error calculations (prevents stoping at local minima)

but slowwww

Large minibatch : provides Computational boost but requires more memory + Computational resources

★ Epochs

Use early stopping

ex : Stop training if the model did not improve in the 10 or 20 steps

★ Hidden Units

we can think of it as a learning capacity.

if too large the model overfits

to avoid this use

* Drop out
* Regularization

★ MLP :

3 hidden layers better than 2

but deeper is not benificial

★ CNN

the deeper the better

RNN hyperparameters :

1. Cell Type:
   - ★ – LSTM
   - – RNN
   - ★ – GRU
2. how deep (Layers).
3. Embedding Dim.
   - * try 50 – 200
   - * then 500 – 1000