Training Neural Networks

Thursday, 27 May 2021 5:22 PM

I. Training Optimization

1. Testing

Testing set helps identify which model between the available model is the best

always choose a simple model over Complex models

Some models overfit or under fit
And so testing data can tell if we are overfitting or not

underfit
Error due to bias

overfit
Error due to variance

2. Early stopping

Stop when both the train and test accuracy are the same or close to each other

avoids under & overfit

3. Regularization

$\sigma(1+1)=0.88$ → $x_1 + x_2$ or $10x_1 + 10x_2$ } Same Line ← smaller Error

$\sigma(10+10)\approx 1$

2 models can give the same split of data but the complex model will try to prove it is better because it has smaller error

large Coefficient ⟶ overfitting

Regularization: penalize large weights in the error function

$$E(w) = -\frac{1}{m}\sum_{i=1}^{m} y_i \ln(\hat{y}_i) + (1-y_i)\ln(1-\hat{y}_i) + \text{Regularization}$$

L1: $\lambda(|w_1|+\ldots+|w_n|)$
- Large if w_i large
- Good for Feature Selection
- $(1,0,0,1,0)$
- Small w_i goes to 0
- Large w_i goes to 1

L2: $\lambda(w_1^2+\ldots+w_n^2)$
- Large if w_i large
- Good for trainning
- $(0.5, 0.3, -0.2, 0.4, 0.1)$
- Keeps vectors homogeneously small

4. Drop out

Turning of neurons in the hidden layer based on probability .

Helps with overfitting and improves training in general

5. local minima

the initial random weight are important & Could lead to local minima

Sol :

* Random restart

6. use Stochastic gradients descent

batch gradient descent :- uses whole data for training

Stochastic gradient descent takes small subset of data instead of the whole

7. Avoid learning Rate decay :-

Too big of learning rate will miss the minimum
Too small of learning rate will take to long to reach minimum
An optimal learning rate will reach minimum within an optimal time

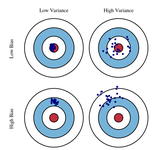
★ Bais vs variance

The error used to evaluate model is composed of
- bias error,
- variance error,
- noise error ← Can't be chaged

} can be changed

Bias Error: Difference between predicted value and expected
model makes assumptions that certain features are not important

Variance Error: Model takes into account noise in data



★ Bertraind Russell

« The whole problem with A.I. is that bad models are so certain of themselves, & good model so full of doubt »

★ Vanishing Gradients

due to sigmoid activation function