Week 3a - Hyperparameter tuning

Friday, August 14, 2020 11:29 PM

① Tuning Process

→ Hyperparameters

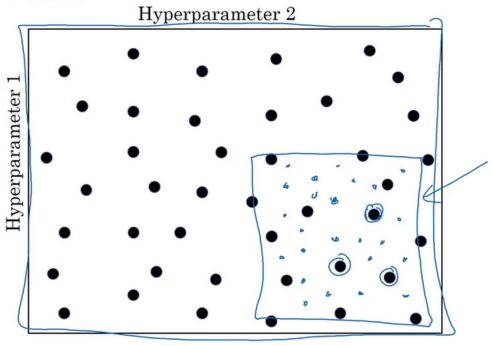
- learning rate α
- β ≈ 0.9
- $\beta_1, \beta_2, \varepsilon$ 10^{-8}
- no. of layers
- no. of hidden units
- learning rate decay
- mini-batch size

→ most important to tune
→ moderate importance
→ least important

→ Try random parameters : Don't use grid

→ Use coarse to fine sampling

Coarse to fine



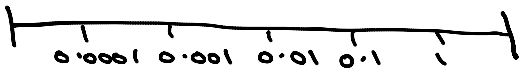
② Using an appropriate scale to pick hyperparameters

→ Picking hyperparameters at random

- For layers we might search 1,...4
- for $n^{[l]}$ → 50...100 might be reasonable

→ appropriate scale for hyperparameters

- $\alpha = 0.0001 \ldots 1$ (small … large)
  - searching at random doesn't work since 90% values are between 0.1 & 1
  - only 10% between 0.0001 & 1
- Hence we use logarithmic scale



- Implementing in python

```
r = -4 * np.random.rand()   → r ∈ [-4, 0]
α = 10^r                    → 10^-4 ... 10^0
```

Generally,

to generate values from 10^a to 10^b logarithmically

Generally,

to generate values from 10^a to 10^b logarithmically

generate r randomly between a & b and 10^r gives req. values

→ Hyperparameters for exponentially weighted averages

$\beta = 0.9 \ldots 0.999$

instead use

$1-\beta = 0.1 \ldots 0.001$ $\quad 10^{-1}$ to 10^{-3}

$\Rightarrow r \in [-3, -1]$

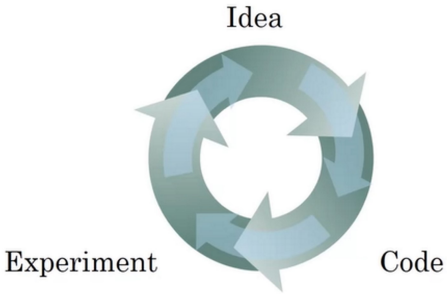
$1-\beta = 10^r$

$\beta = 1 - 10^r$

This is done because β is sensitive to small changes when β is larger (ie) $0.999 \rightarrow 0.9995$ make more diff than $0.9000 \rightarrow 0.9005$ even though difference is same

③ Hyperparameters tuning in practice:

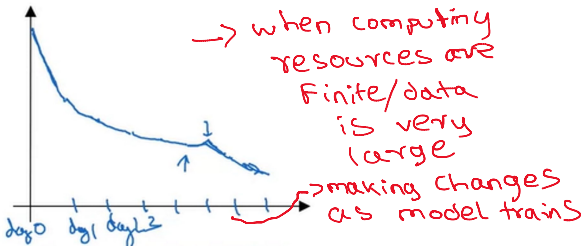
Re-test hyperparameters occasionally



- NLP, Vision, Speech, Ads, logistics, ...
- Intuitions do get stale. Re-evaluate occasionally.

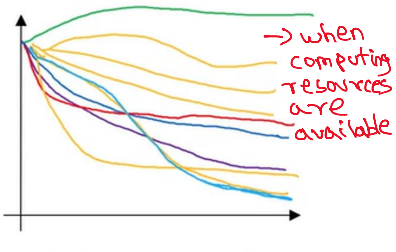
Andrew Ng

Babysitting one model



→ when computing resources are Finite/data is very large → making changes as model trains

Training many models in parallel



→ when computing resources are available