

Chapter 1

Tricks

Real Ensembles

- Train multiple independent models
- At test time average their results

Enjoy $\sim 2\%$ extra performance.

Simulated Ensembles

Fake ensembles. As you are training, normally, you save a checkpoint after each epoch to figure out what was your validation performance at that point. At test time average multiple model checkpoints of a single model. Enjoy $\sim 1\%$ extra performance

Average parameter vector

Keep track of (and use at test time) a running average parameter vector. `params_test` is a running sum exponentially decaying. One way to visualize it is the following. Think in a case of optimizing a bowl like function and you are bouncing around the minimum, then taking the average of all this steps gets you closer to the minimum.

```
while True:
    data_batch = dataset.sample_data_batch()
    loss = network.forward(data_batch)
    dparams = network.backward()
    params += update_params() //instead of using this parameters at test time
    params_test = 0.995*params_test + 0.005*params //use this ones
```

This can give you a small boost.

Use float32 instead of float64

Just that. In fact if you can work with float16 do it.

ReLU bias initialization

It is a good practice to initialize them with a slightly positive initial bias to avoid "dead neurons".