Deep Neural Networks

DNN design - hyperparameters

Architecture: what layers and how (CNN, pooling, RNN, FC, …)
Initialization
# of layers
# of nodes per layer
Activation function

Learning rate
# of iterations
Stopping criteria
Minibatch

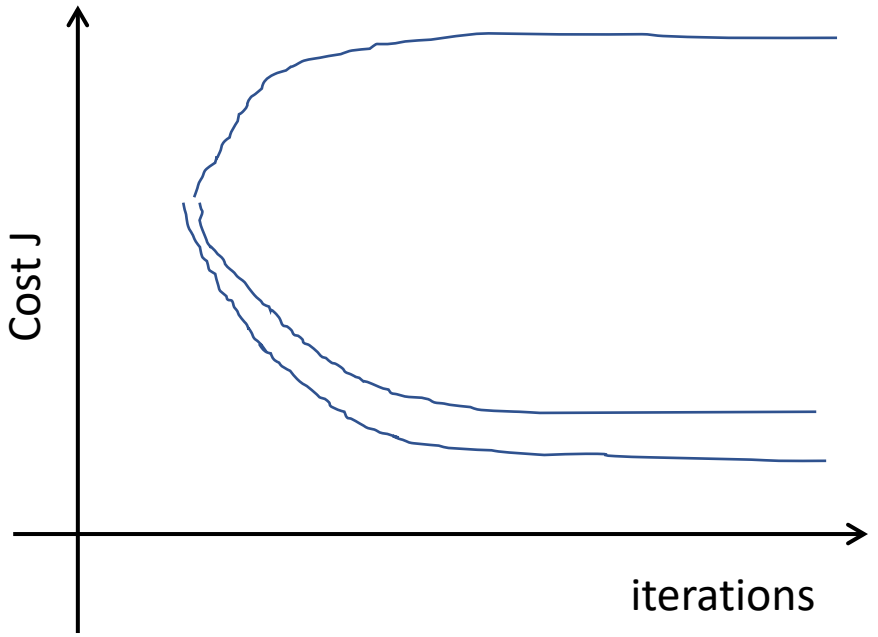
Momentum
Regularization

Transfer learning

It can take days to test a single idea

Designing a deep neural network is an empirical process

Use learning curves to help diagnose!



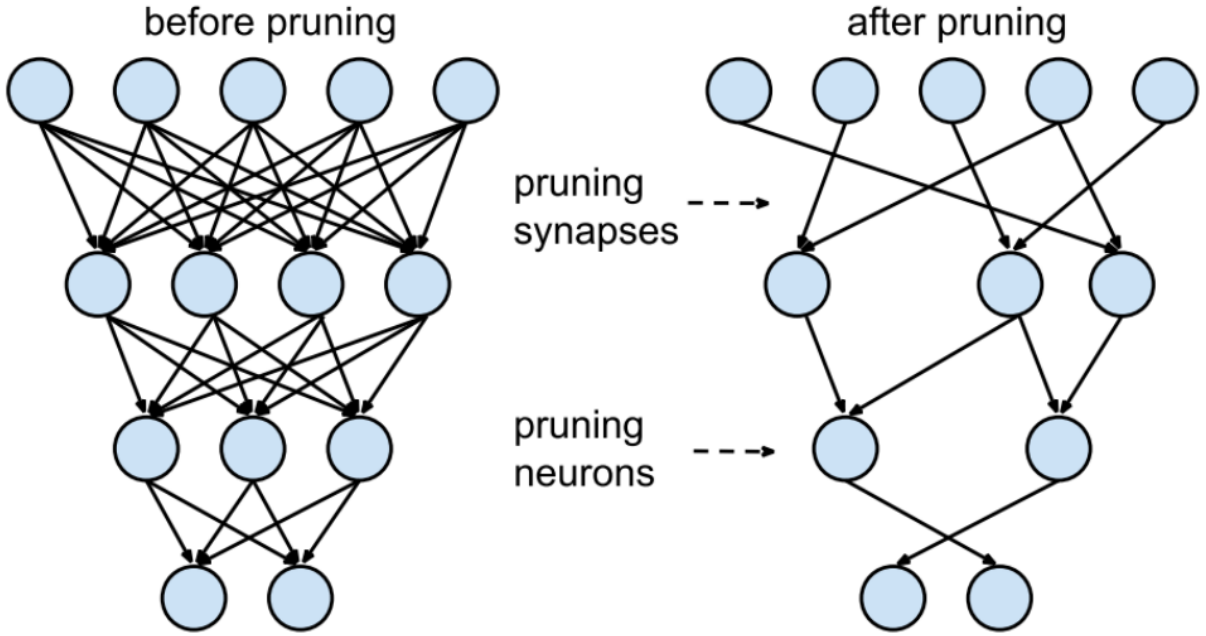
Overfitting

- Good sign that your hypothesis can memorize data, you have a solution!
- Monitor training/validation error cost
- Regularization

Roadblocks: Computation complexity (time), Excessive number of model parameters (overfitting)

Regularization methods (pruning, weight decay, sparsity, early stopping, dropout, data augmentation…)

Pruning



Pruning –
- an iterative process
- Start from a trained model
- Remove the smallest weights iteratively

Stochastic gradient decent –
on its own is self-regularizing, unused weights then to zero over training
Metaphor from the brain, some pathways got strengthened over learning

The overall architecture and properties of each layer impact on how much pruning

For example
ResNet 50 – 90% matches the baseline accuracy on ImageNet dataset
faster inference 4 core CPU 1) X100 ms/image in 2016, 2) X10 ms/image in 2020, 3) ~10 ms/image after pruning)

Dropout

The main idea:

- Randomly selecting nodes to be dropped-out with a given probability (e.g. 20%) each weight update cycle
- To achieve a smaller neural network thus to prevent overfitting
- Randomly eliminate a unit, then downstream neuron cannot rely on any particular input feature, thus encourage spreading out weights, viewed as an adaptive version of the L2 regularization

Implementation:

- Can set different dropout rates and usually a higher rate for layers with too many weights.
- Can treat dropout rate as a hyperparameter and use cross validation to determine an appropriate rate
- When using error cost function J for diagnosis, turn off dropout for evaluation

Computer vision often use dropout