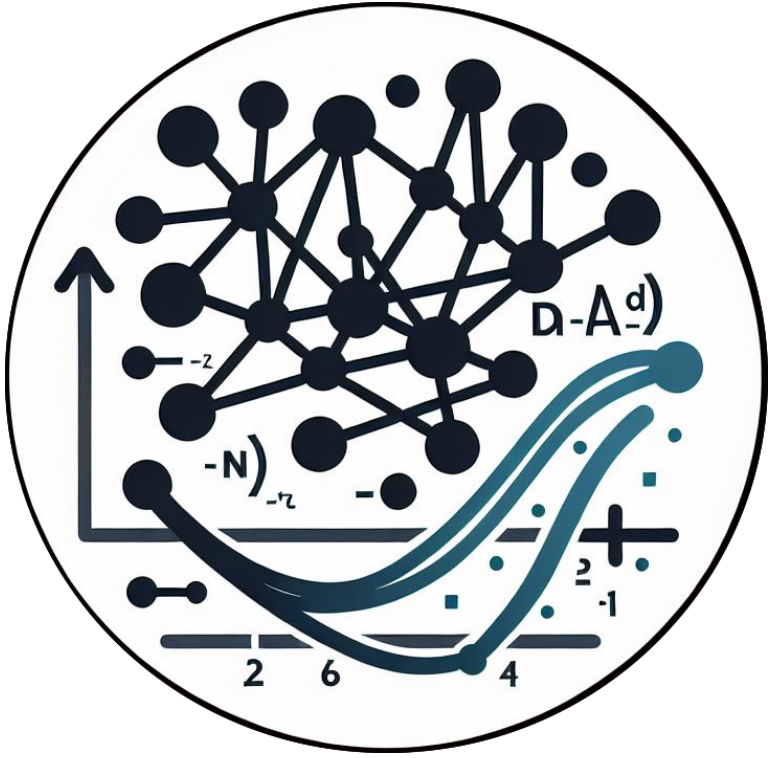
Regularization

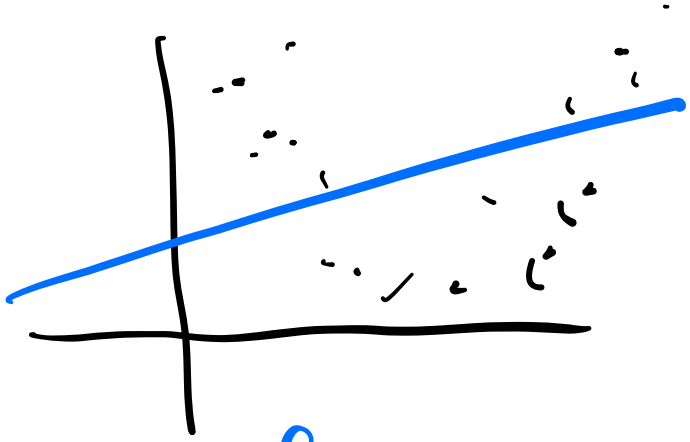


Deep Learning for Engineers

Andrew Ning
aning@byu.edu

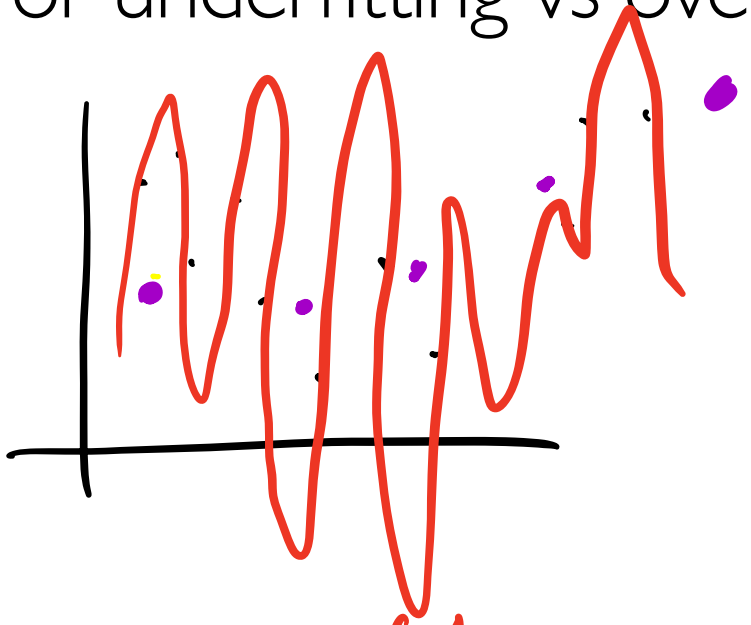
Midterm

Validation (bias vs variance or underfitting vs overfitting)



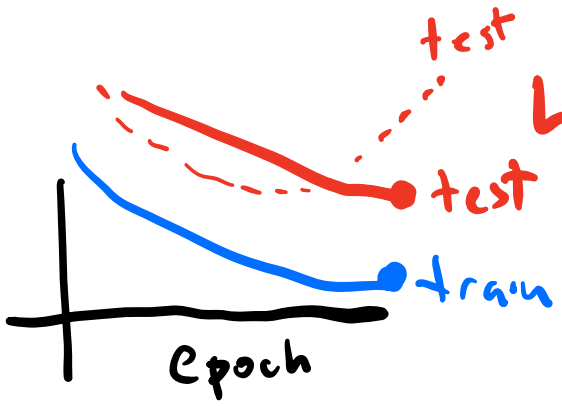
underfit

high bias



overfit

high variance



self.lin1 = nn.Linear(5, 20)

self.lin3 = nn.Linear(20, 50)

x_1 = self.lin1(x)

x_3 = self.lin3($x_1 + x_2$)

First address high bias (underfitting)

Simpler Network

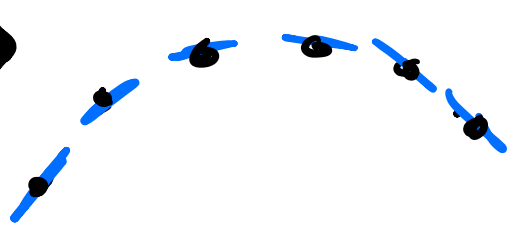
THE LOTTERY TICKET HYPOTHESIS: FINDING SPARSE, TRAINABLE NEURAL NETWORKS

Jonathan Frankle
MIT CSAIL
jfrankle@csail.mit.edu

Michael Carbin
MIT CSAIL
mcarbin@csail.mit.edu

Get more data

- different data.
- impose physics e.g. collocation points
- augment data. →
- synthetic data
- add noise



Add more physics

Early Stopping

Weight regularization

add a loss term

$\gamma \|w\|_2^2$ ← 2 norm

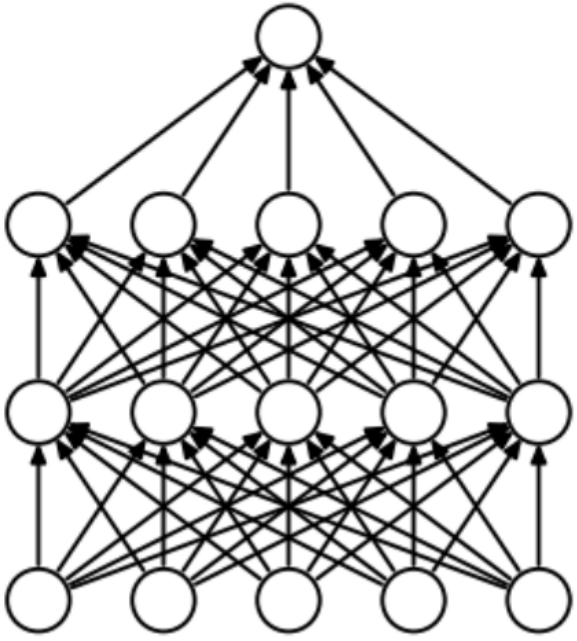
`torch.optim.Adam( ... , weight_decay=1e-4)`

$\gamma \|w\|_1$

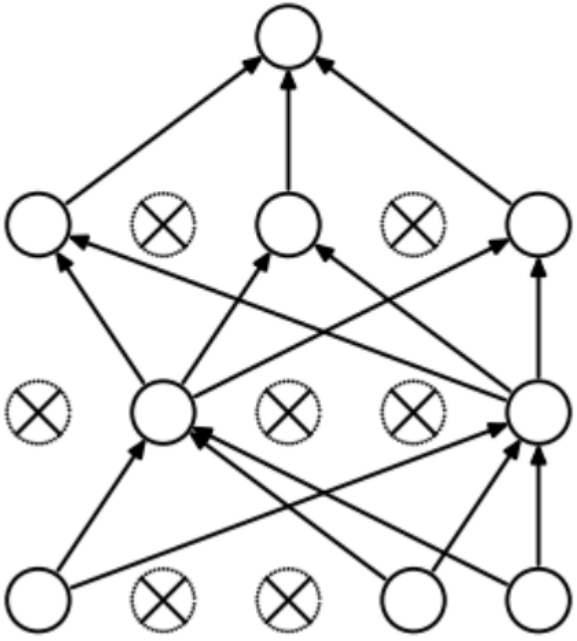
`l1_norm = sum(p.abs().sum() for p in model.parameters())`

Ensemble averaging

Dropout

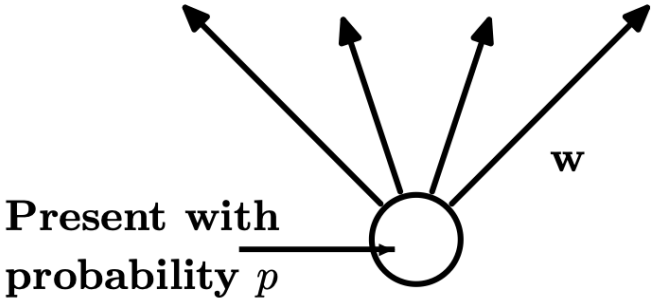


(a) Standard Neural Net

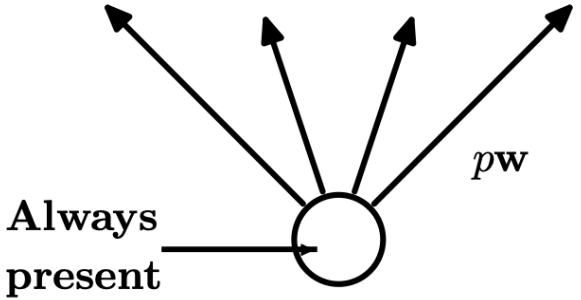


(b) After applying dropout.

Dropout: A Simple Way to Prevent Neural Networks from Overfitting.
Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, Ruslan Salakhutdinov
Journal of Machine Learning Research 15 (2014) 1929-1958



(a) At training time



(b) At test time

```
nn.Sequential(
    ⋮
nn.Dropout(0.2)
)
```

after each activation. (not on inputs)