

# Momentum, Convergence, and Overfitting

## Lesson 8 – Section 5

PROFESSIONAL & CONTINUING EDUCATION  
UNIVERSITY of WASHINGTON



### Momentum

Momentum is a method for speeding up the ANN training process

Can also reduce getting caught in local minima

The tradeoff is speed vs. accuracy

Easy to implement, difficult to find the right value

PROFESSIONAL & CONTINUING EDUCATION  
UNIVERSITY of WASHINGTON

## Convergence

Some neural networks do not “converge”

Often this is a result of incorrectly set hyperparameters

Can also be badly “formatted” input data

PROFESSIONAL & CONTINUING EDUCATION  
UNIVERSITY of WASHINGTON

## Overfitting

### Causes

- More hidden layers can model more complex problems, but can also lead to overfitting
- Assuming the training data is fixed
- Too many iterations (overshooting generalizability)

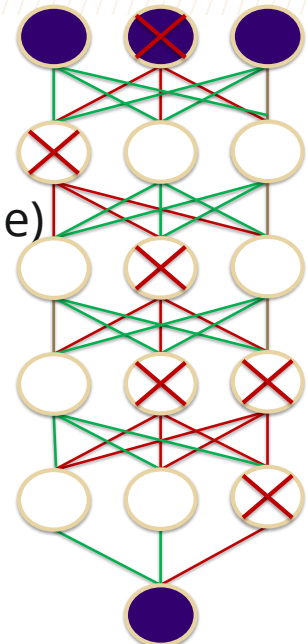
### Prevention

- Stopping early
- More data, randomly sampled
- Cross validation/ensemble neural nets
- Regularization
  - >L1 or L2 regularization
  - >Drop out

PROFESSIONAL & CONTINUING EDUCATION  
UNIVERSITY of WASHINGTON

## Dropout

- Randomly “thin” the network by dropping nodes (hidden and visible) and arcs
- Enables more complex networks with lower cost
- Used for regularization--multiple runs (averaged)



## Pre-Trained Neural Networks

- Starting from scratch to build a neural network
- Reduces time to build a working model
- Accuracy can be either positively or negatively influenced

## Comparison Between SVMs and ANNs

### SVMs

- Deterministic algorithm
- Uses well-known, quadratic programming
- Generalizes well;
- Learns in batches only
- Uses kernels to learn complex functions

### Neural Networks

- Deterministic or stochastic
- Uses learned weights
- Generalizes well
- Can be updated
- Learns complex functions natively

PROFESSIONAL & CONTINUING EDUCATION  
UNIVERSITY of WASHINGTON

## End of Lesson 8

---

### SVM and Neural Nets

PROFESSIONAL & CONTINUING EDUCATION  
UNIVERSITY of WASHINGTON

