CIS 678 Machine Learning

Introduction to Neural Networks (cont.)

Training Neural Network Challenges

- Intractable gradients
  - Vanishing, and
  - Exploding gradients

Training Neural Network Challenges

- Intractable gradients
  - Vanishing, and
  - Exploding gradients
- Various normalizations
  - Input normalization (standard scalar)
  - Batch normalization
  - Layer normalization

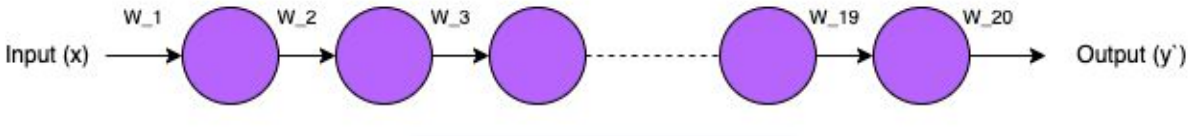
Training Neural Network Challenges

- Intractable gradients
  - Vanishing, and
  - Exploding gradients
- Various normalizations
  - Input normalization (standard scalar)
  - Batch normalization
  - Layer normalization
- Controlling overfitting
  - Regularization
  - Early stopping
  - Drop out

Training Neural Network Challenges

- Intractable gradients
  - Vanishing, and
  - Exploding gradients

Vanishing and Exploding Gradients

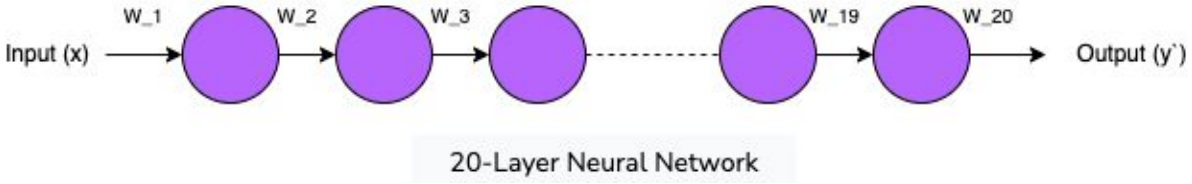


20-Layer Neural Network

$$o_1 = a_1(w_1.x)$$

ref

Vanishing and Exploding Gradients



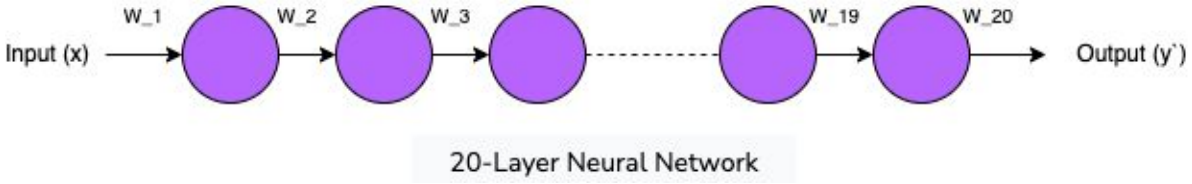
20-Layer Neural Network

$$o_1 = a_1(w_1.x)$$

$$o_2 = a_2(w_2.a_1(w_1.x))$$

ref

Vanishing and Exploding Gradients



$$o_1 = a_1(w_1.x)$$

$$o_2 = a_2(w_2.a_1(w_1.x))$$

..... (l=20)

ref

Whiteboarding

- Intractable gradients
  - Vanishing, and
  - Exploding gradients
- Through normalization
  - Input normalization (standard scalar)
  - Batch normalization
  - Layer normalization
- Controlling overfitting
  - Regularization
  - Early stopping
  - Drop out

Whiteboarding

QA