

Machine Learning

Normalization

Curtis Larsen

Utah Tech University—Computing

Spring 2025

Objectives

Outline

Objectives:

- ▶ Review vanishing/exploding gradient problem
- ▶ Understand and implement batch normalization
- ▶ Understand and implement gradient clipping

Batch Normalization

Idea

- ▶ Vanishing/exploding gradient problems can arise during training.
- ▶ Like `StandardScaler` before/after each layer.
- ▶ Zero centers and normalizes each batch of inputs.

Math

$$\boldsymbol{\mu}_B = \frac{1}{m_B} \sum_{i=1}^{m_B} \mathbf{x}^{(i)} \quad (1)$$

$$\sigma_B^2 = \frac{1}{m_B} \sum_{i=1}^{m_B} (\mathbf{x}^{(i)} - \boldsymbol{\mu}_B)^2 \quad (2)$$

$$\hat{\mathbf{x}}^{(i)} = \frac{\mathbf{x}^{(i)} - \boldsymbol{\mu}_B}{\sqrt{\sigma_B^2 + \epsilon}} \quad (3)$$

$$\mathbf{z}^{(i)} = \boldsymbol{\gamma} \otimes \hat{\mathbf{x}}^{(i)} + \boldsymbol{\beta} \quad (4)$$

Math

$$\boldsymbol{\mu}_B = \frac{1}{m_B} \sum_{i=1}^{m_B} \mathbf{x}^{(i)} \quad (1)$$

$$\sigma_B^2 = \frac{1}{m_B} \sum_{i=1}^{m_B} (\mathbf{x}^{(i)} - \boldsymbol{\mu}_B)^2 \quad (2)$$

$$\hat{\mathbf{x}}^{(i)} = \frac{\mathbf{x}^{(i)} - \boldsymbol{\mu}_B}{\sqrt{\sigma_B^2 + \epsilon}} \quad (3)$$

$$\mathbf{z}^{(i)} = \boldsymbol{\gamma} \otimes \hat{\mathbf{x}}^{(i)} + \boldsymbol{\beta} \quad (4)$$

$\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ are learnable parameters. $\boldsymbol{\mu}$ and σ are estimated during training, and used during prediction.

Implementation

```
model.add(keras.layers.BatchNormalization())  
model.add(keras.layers.Dense(units, ...))
```


Gradient Clipping

Ideas

- ▶ Exploding gradients cause the fit to diverge.
- ▶ Limiting the gradient magnitude at each step is called gradient clipping.
- ▶ Clipping individual components in the gradient.
- ▶ Clipping the gradient magnitude.
- ▶ Can be used as an alternative to Batch Normalization, or in addition to Batch Normalization.

Implementation

```
keras.optimizers.SGD(clipvalue=1.0)
# OR
keras.optimizers.SGD(clipnorm=1.0)
```

Summary

Topics

- ▶ Batch normalization
- ▶ Gradient clipping