Vanishing and Exploding Gradient

Fay Elhassan, Vongai Mitchell, Solafa Fadlallah, Bahati Kilongo

African Masters of Machine Intelligence/AIMS-Senegal

Course: Foundation of machine learning, taught by Prof. Moustapha Cisse

April 19, 2023

AIMS African Institute for Mathematical Sciences SENEGAL

Overview

Our presentation will focus on four points which are:

1. Neural Network
2. Vanishing and Exploding gradient
   - Identification
   - Causes
   - Consequences
   - Solutions
     - Architectures that are designed to deal with Vanishing/Exploding Gradient
3. Python code
4. Conclusion

AIMS African Institute for Mathematical Sciences SENEGAL

Simple Neural network

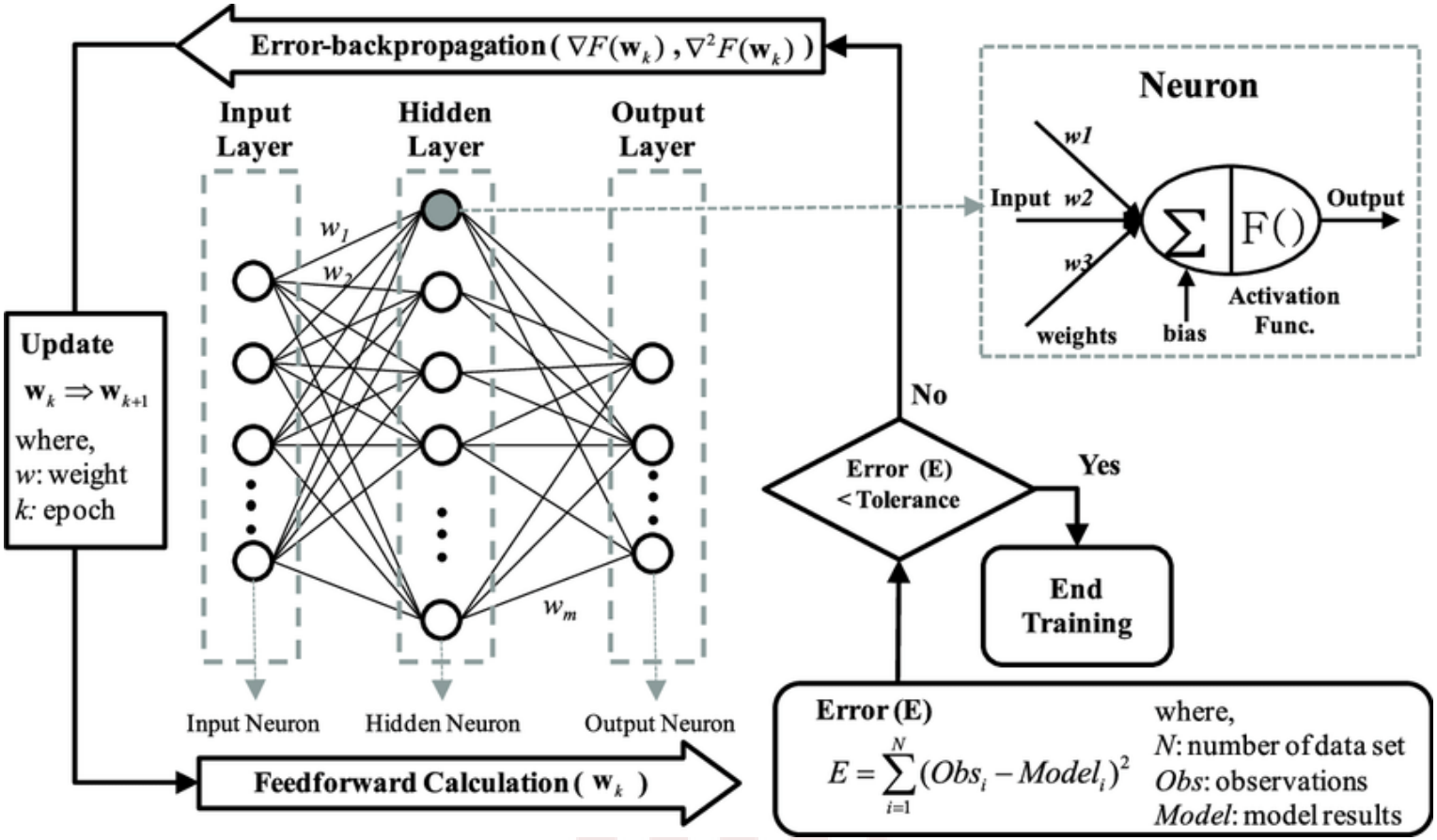

Figure: Simple Neural Network.Ref:researchgate

AIMS African Institute for Mathematical Sciences SENEGAL

Vanishing/ Exploding gradient

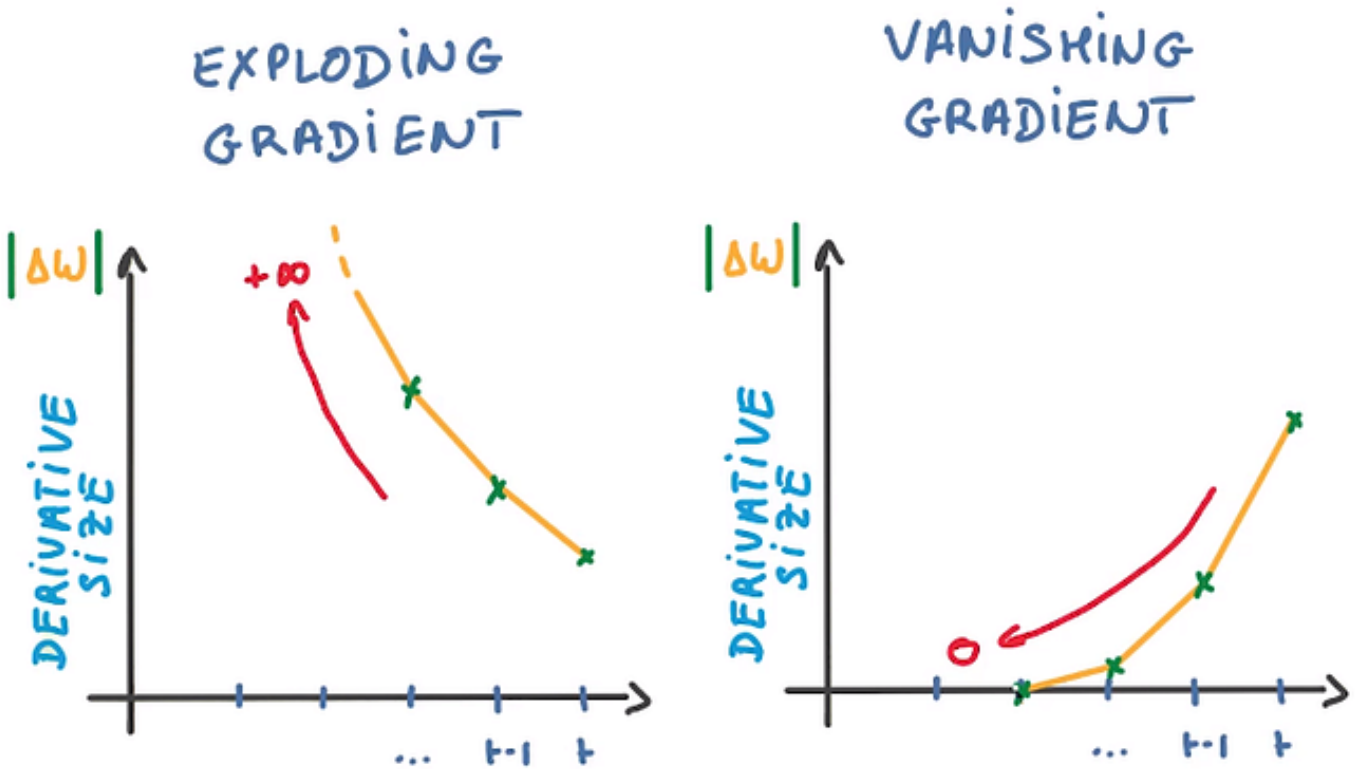

Figure: Vanishing/ Exploding gradient. ref:comet

Overview

1. Neural Network
2. Vanishing and Exploding gradient
   - Identification
   - Causes
   - Consequences
   - Solutions
     - Architectures that are designed to deal with Vanishing/Exploding Gradient
3. Python code
4. Conclusion

How to identify vanishing/exploding gradient ?

Vanishing gradient is identified:

- Parameters of higher layers change significantly while those of lower layers do not change much.

How to identify vanishing/exploding gradient ?

Vanishing gradient is identified:

- Parameters of higher layers change significantly while those of lower layers do not change much.

Exploding gradient is identified:

- Parameters experience exponential growth.
- Model weights may become NaN very quickly during training.

AIMS African Institute for Mathematical Sciences SENEGAL

Overview

1. Neural Network
2. Vanishing and Exploding gradient
   - Identification
   - Causes
   - Consequences
   - Solutions
     - Architectures that are designed to deal with Vanishing/Exploding Gradient
3. Python code
4. Conclusion

Why does the gradient even vanish/explode ?

- Very small/ large losses due to the initial weights.

Why does the gradient even vanish/explode ?

- Very small/ large losses due to the initial weights.
- Derivative closed to zero.

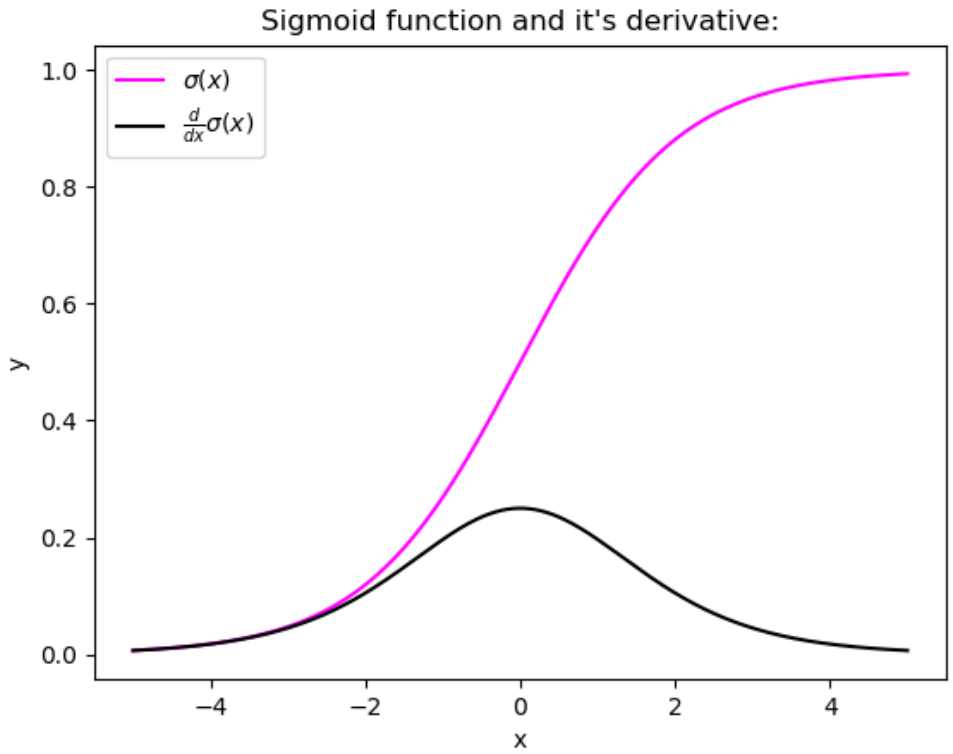

Figure: Derivative of Sigmoid. ref:comet

AIMS African Institute for Mathematical Sciences SENEGAL

Overview

1. Neural Network
2. Vanishing and Exploding gradient
   - Identification
   - Causes
   - Consequences
   - Solutions
     - Architectures that are designed to deal with Vanishing/Exploding Gradient
3. Python code
4. Conclusion

AIMS African Institute for Mathematical Sciences SENEGAL

How does Vanishing/Exploding gradient affect the training of the model ?

Vanishing gradient

- The model learns slowly and training might even stagnate.
- Model performance is poor.

How does Vanishing/Exploding gradient affect the training of the model ?

Vanishing gradient

- The model learns slowly and training might even stagnate.
- Model performance is poor.

Exploding gradient

- The model loss can become NaN and the model experiences avalanche learning.
- Model performance is poor.

Overview

1. Neural Network
2. Vanishing and Exploding gradient
   - Identification
   - Causes
   - Consequences
   - Solutions
     - Architectures that are designed to deal with Vanishing/Exploding Gradient
3. Python code
4. Conclusion

How to resolve Vanishing/Exploding gradient issues ?

Table of techniques and architectures

Some architectures and techniques deal with both vanishing and exploding gradient issues, other techniques deal only with vanishing gradient issues, and others deal only with exploding gradient issues.

Vanishing	**Exploding**	**Vanishing-Exploding**
Change activation functions	L2 norm regularization	Weight initialization
Reducing model complexity	Gradient clipping	Batch normalization
Using a better optimizer		Some Architectures: ResNets,LSTM,GRU

Resnet

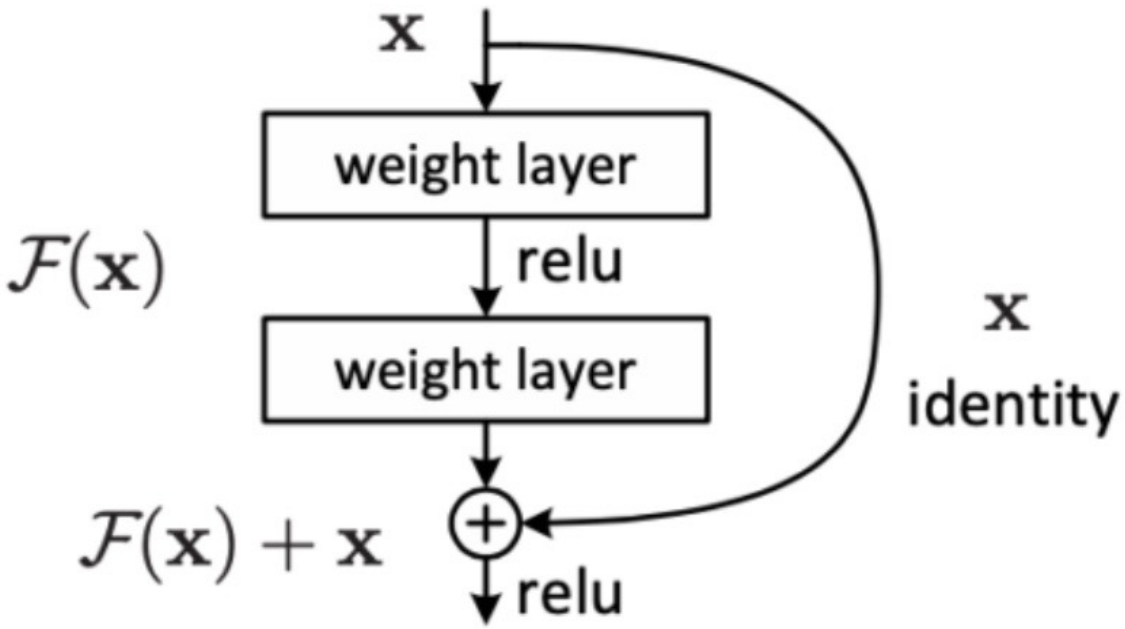

Figure: Residual Neural Networks. (Ref:Deep Residual Learning)

LSTM

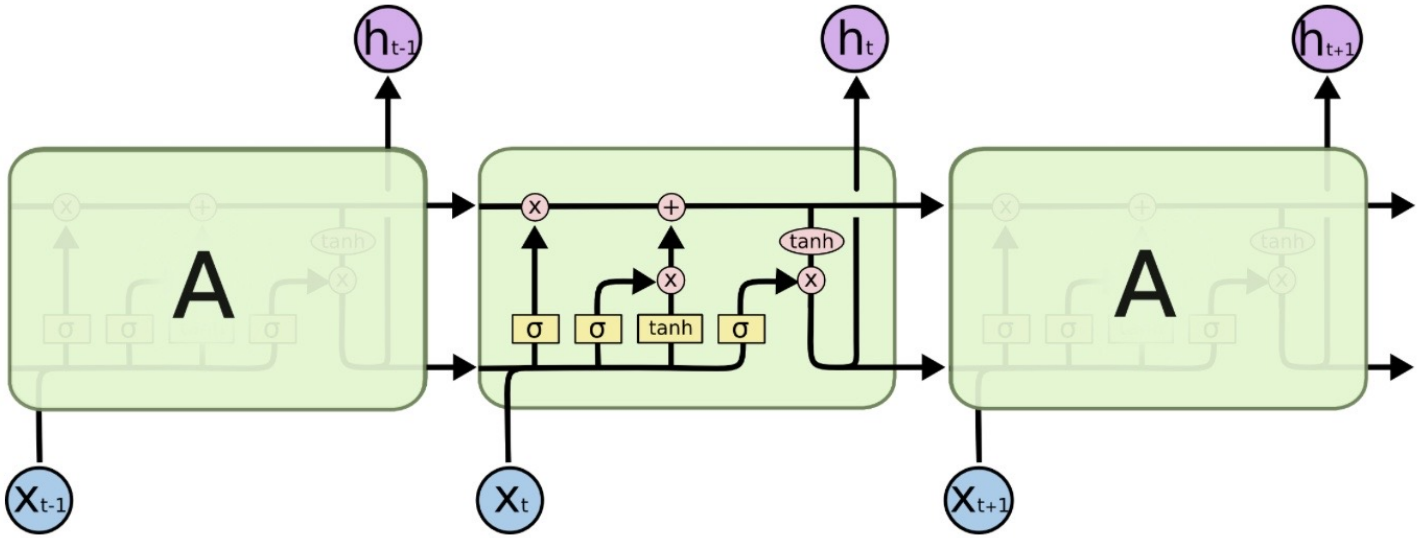

Figure: Long-Short Term Memory. (Source: Research Gate)

Python

This is our Python code. Click here to visit the notebook.

Conclusion

Vanishing and exploding gradients are a major problem in neural networks, however the techniques explained in this presentation have been used in addressing these challenges, training a deep neural network without these techniques is like trying to balance a pencil upright on your finger.

[1] Xavier Glorot and Yoshua Bengio. "Understanding the difficulty of training deep feedforward neural networks". In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. Ed. by Yee Whye Teh and Mike Titterington. Vol. 9. Proceedings of Machine Learning Research. Chia Laguna Resort, Sardinia, Italy: PMLR, 13–15 May 2010, pp. 249–256. URL: https://proceedings.mlr.press/v9/glorot10a.html.

[2] Kaiming He et al. *Deep Residual Learning for Image Recognition*. 2015. arXiv: 1512.03385V1 [cs.CV].

[3] Ismail San Kerem SARI İsmail Hakkı İPEK. *TradeNIC: Trading-specific SmartNIC Design for Low Latency and High Throughput Algorithmic Trading*. 2022.

[4] Yuki Tatsunami and Masato Taki. *Sequencer: Deep LSTM for Image Classification*. 2023. arXiv: 2205.01972v4 [cs.CV].