

Natural Language Processing (NLP)

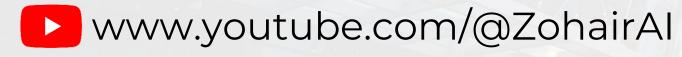
Vanishing and Exploding Gradients

Equipping You with Research Depth andIndustry Skills

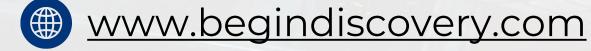
By:

Dr. Zohair Ahmed











Vanishing and Exploding Gradients

- **Problem Overview**: The vanishing and exploding gradient problems are issues encountered during neural network training, particularly in deep networks and RNNs.
- **Goal**: Understand how these problems affect training and how they impact the learning process in neural networks and RNNs.

Vanishing Gradient Problem

- Scenario: We have a neural network designed to predict if a person will buy insurance based on factors like age, education, and income.
- During training, the network goes through a **forward pass** to calculate the loss and then a backward pass to adjust weights.
- **Key Process**: In the backward pass, gradients are computed to adjust weights using backpropagation.

www.youtube.com/@ZohairAl

The gradient tells us how much the weights need to change to reduce the loss.

How Gradients Affect Weight Updates

Gradient Calculation:

- Gradients are the derivative of the loss with respect to each weight, indicating how the loss changes as the weight changes.
- If the gradient is small, the weights change by a very small amount, slowing down the learning process.
- Problem:
- When gradients are very small (close to zero), weight updates become insignificant, causing the network to **learn very slowly**.
- This is known as the vanishing gradient problem.

Vanishing Gradients in Neural Networks

- Impact of Deep Networks:
- As the number of layers increases, the gradients are multiplied through each layer during backpropagation.
- If the gradients in earlier layers are small, they become **even smaller** as they are propagated backward.
- This causes the weights in the earlier layers to change very little, impeding the network's ability to learn effectively.
- Example:
- With a **deep neural network** (many layers), the effect of the **initial inputs** diminishes as you move through the layers, resulting in slow or no learning in the earlier layers.

Exploding Gradient Problem

- Scenario: If the derivatives (gradients) are too large, the gradient becomes exploded, leading to extremely large weight updates.
- This causes the weights to grow rapidly and the model to diverge, making learning unstable.
- Impact: In such cases, the neural network might fail to converge and lead to oscillations or overly large updates, often resulting in a NaN (Not a Number) error in training.

Vanishing Gradient in RNNs

- Use Case: RNNs in Natural Language Processing (NLP)
- Example: **Text Autocompletion** in Gmail:
- Input: "Today due to my current job situation and family condition, I..."
- Autocompleted Output: "... need to take a loan."
- Problem in RNNs:
- In RNNs, earlier words in the sequence (e.g., "Today") should have an impact on the later words (e.g., "need").
- However, due to the vanishing gradient problem, the influence of earlier words diminishes as you
 move through the sequence.
- **Effect**: As the sequence length increases, the influence of earlier words **vanishes**. This leads to difficulty in learning long-range dependencies in sequences.

Backpropagation and Vanishing Gradients in RNNs

- Backpropagation in RNNs:
- As the error is backpropagated through the time steps in RNNs, the gradient becomes progressively smaller.
- Earlier time steps (like "Today") have a very small influence on the weight updates in later time steps.
- Impact:
- RNNs suffer from short-term memory, where long-range dependencies are difficult to learn due to vanishing gradients.

Vanishing Gradient Example in RNNs

- Autocompletion Example:
- Input: "Today due to my current job situation..."
- Expected output: "... I need to take a loan."
- Effect of Vanishing Gradients:

By: Zohair Ahmed

- The word "Today" has a significant impact on the output, but as the sentence grows, the influence of "Today" becomes weaker and weaker.
- By the time the network reaches the word "need", the influence of the initial word ("Today") is almost lost.

Solutions to Vanishing and Exploding Gradients

- Vanishing Gradient Solutions:
- LSTM (Long Short-Term Memory): A special type of RNN designed to combat the vanishing gradient problem by using gates to control the flow of information.
- GRU (Gated Recurrent Unit): A simpler variant of LSTM that also addresses the issue of vanishing gradients by using a gating mechanism.
- Exploding Gradient Solutions:
- Gradient Clipping: If gradients exceed a certain threshold, they are scaled down to prevent the weights from exploding.
- Weight Regularization: Regularizing the weights during training can help prevent the gradients from becoming too large.

Solutions to Vanishing and Exploding Gradients

- Vanishing Gradient Solutions:
- LSTM (Long Short-Term Memory): A special type of RNN designed to combat the vanishing gradient problem by using gates to control the flow of information.
- GRU (Gated Recurrent Unit): A simpler variant of LSTM that also addresses the issue of vanishing gradients by using a gating mechanism.
- Exploding Gradient Solutions:
- Gradient Clipping: If gradients exceed a certain threshold, they are scaled down to prevent the weights from exploding.
- Weight Regularization: Regularizing the weights during training can help prevent the gradients from becoming too large.