5 - Deep Learning Overview

5.7 - Common Problems in Deep Learning

Wendley S. Silva

Setembro, 2020

Index

Data Imbalance (1)

- Problem description: In the dataset consisting of various task categories, the number of samples varies greatly from one category to another. One or more categories in the predicted categories contain very few samples.
- For example, in an image recognition experiment, more than 2,000 categories among a total of 4251 training images contain just one image each. Some of the others have 2-5 images.
- Impacts:
  - Due to the unbalanced number of samples, we cannot get the optimal real-time result because model/algorithm never examines categories with very few samples adequately.
  - Since few observation objects may not be representative for a class, we may fail to obtain adequate samples for verification and test.

HUAWEI

Data Imbalance (2)

Random undersampling

- Deleting redundant samples in a category

Random oversampling

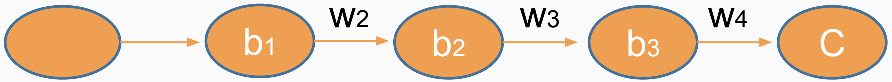
- Copying samples

Synthetic Minority Oversampling Technique

- Sampling
- Merging samples

Vanishing Gradient and Exploding Gradient Problem

- Vanishing gradient: As network layers increase, the derivative value of backpropagation decreases, which causes a vanishing gradient problem.
- Exploding gradient: As network layers increase, the derivative value of backpropagation increases, which causes an exploding gradient problem.
- Cause: $y_i = \sigma_{(z_i)} = \sigma(w_i x_i + b_i)$ Where σ is sigmoid function.

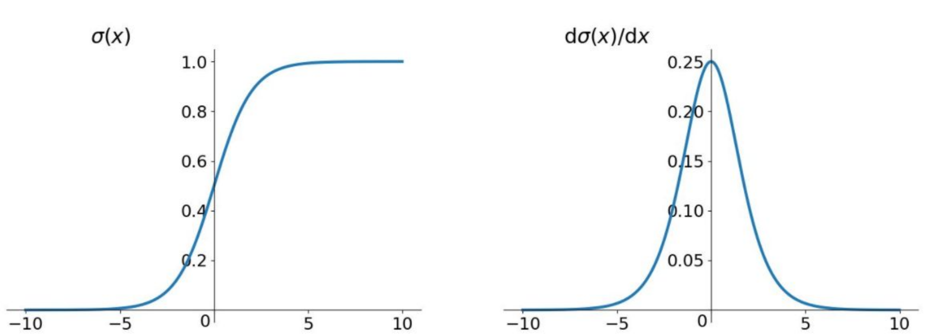


- Backpropagation can be deduced as follows:

$$\frac{\partial C}{\partial b_1} = \frac{\partial C}{\partial y_4}\frac{\partial y_4}{\partial z_4}\frac{\partial z_4}{\partial x_4}\frac{\partial x_4}{\partial z_3}\frac{\partial z_3}{\partial x_3}\frac{\partial x_3}{\partial z_2}\frac{\partial z_2}{\partial x_2}\frac{\partial x_2}{\partial z_1}\frac{\partial z_1}{\partial b_1}$$
$$= \frac{\partial C}{\partial y_4}\sigma'(z_4)w_4\sigma'(z_3)w_3\sigma'(z_2)w_2\sigma'(z_1)x$$

Vanishing Gradient and Exploding Gradient Problem (2)

- The maximum value of $\sigma'(x)$ is $\frac{1}{4}$:



- However, the network weight $|w|$ is usually smaller than 1. Therefore, $|\sigma'(z)w| \leq \frac{1}{4}$. According to the chain rule, as layers increase, the derivation result $\frac{\partial C}{\partial b_1}$ decreases, resulting in the vanishing gradient problem.
- When the network weight $|w|$ is large, resulting in $|\sigma'(z)w| > 1$, the exploding gradient problem occurs.
- Solution: For example, gradient clipping is used to alleviate the exploding gradient problem, ReLU activation function and LSTM are used to alleviate the vanishing gradient problem.

HUAWEI

Overfitting

- Problem description: The model performs well in the training set, but badly in the test set.
- Root cause: There are too many feature dimensions, model assumptions, and parameters, too much noise, but very few training data. As a result, the fitting function perfectly predicts the training set, while the prediction result of the test set of new data is poor. Training data is overfitted without considering generalization capabilities.
- Solution: For example, data augmentation, regularization, early stopping, and dropout

HUAWEI

Thank You!

Next: 6.0 - Development Environments for Deep Learning

Wendley Silva

Setembro, 2020