Getting Started

Prof. Eric A. Suess

March 25, 2020
In Chapter 2 the network architecture, compilation and fitting steps are discussed. The first examples are of feedforward neural networks.

In the network architecture the `layer_dense()` and the activation function is used.

In the compilation steps the loss, the optimizer, and the metric is specified.
Tensors

multidimensional arrays

- 1D
- 2D
- 3D
- 4D
Examples of Tensors

- **Vector data** - 2D tensors of shape \((\text{samples}, \text{features})\)
- **Timeseries data or sequence data** - 3D tensors of shape \((\text{samples}, \text{timesteps}, \text{features})\)
- **Images** - 4D tensors of shape \((\text{samples}, \text{height}, \text{width}, \text{channels})\) or \((\text{samples}, \text{channels}, \text{height}, \text{width})\)
- **Video** - 5D tensors of shape \((\text{samples}, \text{frames}, \text{height}, \text{width}, \text{channels})\) or \((\text{samples}, \text{frames}, \text{channels}, \text{height}, \text{width})\)
Array reshaping

In R **columns** are the primary way to look at arrays.

In Tensorflow **rows** are primary and the `array_reshape()` is used to prepare data for use with keras/tensorflow.
A geometric interpretation of deep learning

Be sure to read this section. The description of what neural networks try to do is very interesting.
Derivatives in Calculus measure the rate of change of a function. In multidimensional space derivatives are called gradients. Gradients are used to minimize/maximize functions.
Stochastic Gradient Decent

Four step process.

1. Draw a batch of training samples $x$ and corresponding targets $y$.
2. Run the network on $x$ to obtain predictions $y_{\text{pred}}$.
3. Compute the loss of the network on the batch, a measure of the mismatch between $y_{\text{pred}}$ and $y$.
4. Compute the gradient of the loss with regard to the network’s parameters (a backward pass).
5. Move the parameters a little in the opposite direction from the gradient—for example, $W = W - (\text{step} \times \text{gradient})$—thus reducing the loss on the batch a bit.
A neural network is composed of multiple tensor operations.

The backpropagation algorithm used the Chain Rule to implement the algorithm.
Examples in Chapter 3

Try to get the examples from Chapter 3 to run.

1. Binary classification, movie reviews

The data preparation uses the Bag-of-Words model. All words in the reviews for the Dictionary or Bag-of-Words. Each word is given a number value.

2. Multi-class classification, newswires

3. Regression, predicting house prices.