

Chapter 1

Other definitions

1.1 Embedding

An embedding maps an input representation, such as a word or sentence, into a vector. A popular type of embedding are word embeddings such as word2vec or GloVe. We can also embed sentences, paragraphs or images. For example, by mapping images and their textual descriptions into a common embedding space and minimizing the distance between them, we can match labels with images. Embeddings can be learned explicitly, such as in word2vec, or as part of a supervised task, such as Sentiment Analysis. Often, the input layer of a network is initialized with pre-trained embeddings, which are then fine-tuned to the task at hand.

1.2 Gradient Clipping

Gradient Clipping is a technique to prevent exploding gradients in very deep networks, typically Recurrent Neural Networks. There exist various ways to perform gradient clipping, but the a common one is to normalize the gradients of a parameter vector when its L2 norm exceeds a certain threshold according to `new_gradients = gradients * threshold / l2_norm(gradients)`.

1.3 Vanishing Gradient Problem

The vanishing gradient problem arises in very deep Neural Networks, typically Recurrent Neural Networks, that use activation functions whose gradients tend to be small (in the range of 0 from 1). Because these small gradients are multiplied during backpropagation, they tend to vanish throughout the layers, preventing the network from learning long-range dependencies. Common ways to counter this problem is to use activation functions like ReLUs that do not suffer from small gradients, or use architectures like LSTMs that explicitly combat vanishing gradients. The opposite of this problem is called the exploding gradient problem.

