

CS224n Lecture-4

We should retrain word2vec vectors from scratch only if we have a huge dataset. We might fine-tune the vectors if we have a sufficiently large dataset, in order of millions of words.

The lecture then discusses the normal forward propagation and backward propagation algorithms. Quick notes could be the use of the error signal for efficiently calculating the gradients. You can use gradient checking for checking your backpropagation algorithm implementation.

Generic stuff, like regularisation, vectorization, non-linearities, optimizers and learning rates were discussed.