CS224n Lecture 3

A significant difference to note between regular machine learning and what we did in word2vec is representational learning. In representation learning, we learn both the weights and the word vectors.
The lecture proceeded to named entity recognition, where we want to assign each word a label like a place, person etc.

A straightforward idea is for every word do the softmax classification. This does not work well because we are neglecting the all-important context information. Next thing we can do is to take a window of vectors and average their word senses and make a prediction for it. This also doesn't work well as it loses the positional information.

So what we do is we stack up the context and the word in a significant vector. And add a hidden layer to it. We want to predict a high score when the target is between the context and low score when it is not between the window.
Finally, the multivariate calculus was discussed.