

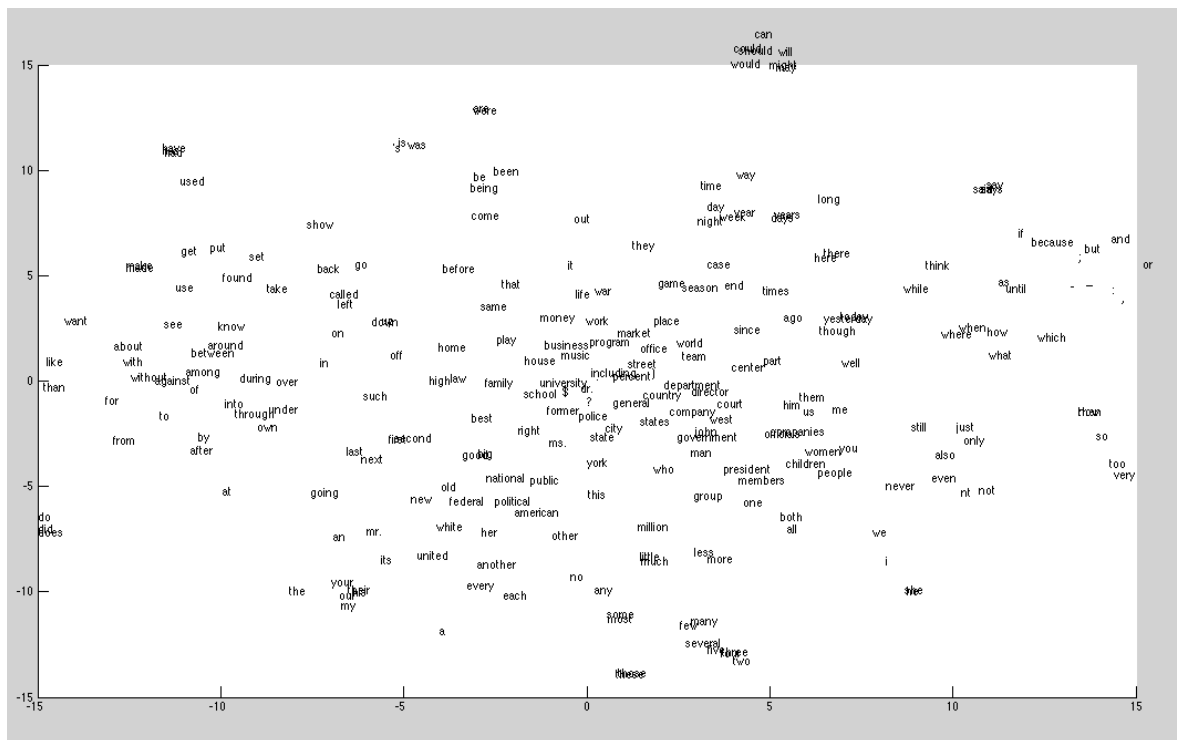
## CC321 Assignment One

### 1) Results with different hyper parameters

Model parameters	Training CE error	Validation CE error	Testing CE error	Number of Epochs
D=8 Num_hid=64	2.854	2.882	2.882	6
D=8 Num_hid=256	2.681	2.745	2.744	8
D=32 Num_hid=64	2.707	2.748	2.756	7
D=32 Num_hid=256	2.530	2.626	2.633	9

Analysis: Increasing the number of hidden units and the number of dimensions in the distributed representation; decreases the cross entropy error on the testing data. However, increasing the network complexity will increase the risk of over fitting the network. This is avoided by the early stopping mechanism implemented.

### 2) Model 2-D distributed representation



### 3) Closest Words

Word	First Nearest	Second Nearest	Third Nearest	Fourth Nearest
may	might	will	also	should
Game	Season	Office	Team	Market
million	former	percent	center	including
against	among	between	without	under

Analysis: Many of the words in the learned model that are close to each other, are actually synonyms in the English language. However, this is not true with most of the words since a word like “million” shouldn’t be close to “former” or “center”. The learned model overall seems to be accurate when compared to the English language.

### 4) Testing the Model

Sentence	First	Second	Third	Fourth
Family also used	to	it	one	them
American political federal	.	days	people	times
Government of United	own	states	.	said
President found the	same	world	game	country

Analysis: The model shows semantics and grammar understanding of the sentences. Some of the words predicted don’t make sense like the “government of united own”. The model performance is better for frequently used words like “found the..” or “also used ..”, but inaccuracies are observed when using the model with less frequent data in the training set like “united”. Punctuations were observed with many predicted words, which also prove the last argument about the word frequencies in the training data effect.

### 5) Distances between words

- (back, go)-> 3.8078
- (back, left)-> 2.4856
- (back, calling)-> 2.5419
- (money, million)-> 2.8199
- (money, music)-> 2.3498
- (right, left)-> 4.8674
- (up, down)-> 2.4236

Analysis: (money, million) should be closer than (money, music); this could be because of the sampling error in the dataset and the lack of some of the words/events in our training set. The same applies for (right, left) and (up, down); they should have the same distance between them.