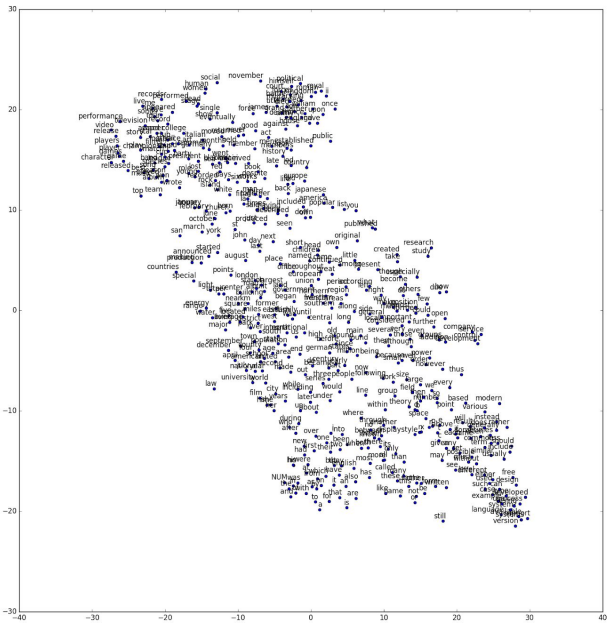
1. Part 1
    a. Skipgram: with embedding size 128. The flow is basically from the (inputs, labels) to the positive entropy loss (posEnt) and negative entropy loss (negEnt). To calculate posEnt, one just needs to calculate posLog = inputs*pos_outputs and then forward it to sigmoid cross entropy with positive labels. Similarly, to negEnt, one first calculates negLog = inputs*neg_outputs ==> sigmoid cross entropy with negative labels. While pos_outputs are lookups of vector 'labels' on embedding tables, neg_outputs are lookups of vector 'sampled_indices' which have to be determined by a candidate sampling function.
    b. I didn't have time and computing resource to fine tune the params. But I find the best params for me is: {'skip_win': 1, 'neg_sampling': 64, 'lr': 0.1, 'dim_vec': 128}
2. Part 2
    a. Some word pairs are very close: (research, study) (september, december) (japanese, america, europe)



    b. Some of them make sense, some not quite:
    Nearest to german: century, french, found, english, became,
    Nearest to most: between, when, more, known, been,
    Nearest to general: left, part, well, states, according,
    Nearest to food: plant, animal, lead, heart, discovered,
    Nearest to cat: shark, eat, juvenile, trek, bird,
    Nearest to eat: cat, bones, breed, skull, shy,
    Nearest to teach: self, doctorate, gown, childs, jung,

3. Part 3
    a. I sum together the embeddings of every word in the document and then divide by the number of words. Together with the label, I then forward to a logistic regression.
    b. Accuracy: 0.849
    Precision:0.845
    Recall:0.849
    F1: 0.8436