HU Extension   Assignment 02   csci-s89A DL for NLP

Handed out: 07/03/2019   Due by 11:59 PM EST on Tuesday, 07/09/2019

Problem 1. Examine examples of analogic reasoning we demonstrated in Lecture 03. Jupyter notebook with those examples is uploaded in the folder for Lecture 03 of the class site. One such example is "what is to Russia, what Paris is to France?". Those four words (Russia, France, Paris and Moscow) should present a polygon with four edges, perhaps a romb or rectangle. Create three more similar analogies and present them in the same PCA plane. We are just curious whether the geometric shapes of those examples are identical or very similar one to another. Please select analogies of very similar nature: countries vs. capitals, people vs food, etc. Do this using the 100-dimensional Glove vectors transformed into Word2Vec format. Use Gensim API. If you are familiar with Spacy or some other NLP API, please be free to use it.
(20%)

Problem 2. Repeat the above experiment with 300-dimensional Glove vectors transformed into Word2Vec format. We are curious whether the shape of above geometric shapes are preserved or modified in the higher dimensional space. Use PCA to make the projections.
(20%)

Problem 3. Repeat the above experiment from problem 1 using Sci-kit Learn T-SNE projections, rather than PCA.
(10%)

Problem 4. Try to use the simplest architecture we discussed in Lecture 4 with two Dense layers to create a de-noising autoencoder. Use the latent space of dimension 64. Report on your findings.
(25%)

Problem 5. Consider image denoising auto decoder described on slide 54 of the notes for Lecture 4. Make an experiment by reducing the number of filters (channels) contained in all Conv2D layers from 32 to 16. Compare results of autoencoders with 32 and 16 channels visually. Could you come up with a technique to compare the quality of denoising more accurately. In either case present the effect of removing noise from handwritten digits 3, 6, and 8.
(25%)

INSTRUCTION:

Please, describe every step of your work and provide intermediate and final results of all of your calculations. Please organize your submission in one Jupyter notebook, if possible. Please, always submit a PDF version of your Jupyter notebook. For issues and comments visit the class Piazza site.