SYLLABUS - MSDS 16:954:577 (Statistical Software)

Brief Python Review : NumPy and Pandas
      creating NumPy arrays, one-dimensional, two-dimensional, arange
      Array attributes, indexing, slicing, reshaping and newaxis, splitting
      Universal functions, aggregation, broadcasting
      Boolean arrays, bitwise logical operators, masking, Array sorting
      Series object, indices of a series object, vectorized operations, Vectorized string operations

Natural Language Processing

Text Processing (Regular Expressions, Tokenizing, Stemming and Lemmatization, Minimum Edit
      Distance Algorithm, Bag of Words, N-grams)

N-gram Language Models: Using n-grams for language modeling , evaluating language models and perplexity, Unknown words, Laplace Smoothing, Backoff and Interpolation

Naive Bayes Classification (Naive Bayes Rule for documents, Laplace smoothing, Binary & Multinomial classifiers, Evaluation Measures: Precision & Recall, F-measure, Confusion Matrix, , Micro and Macro Averaging)

Logistic Regression: Classification using Logistic Regression, Sigmoid and Softmax functions, Learning in Logistic Regression and Gradient Descent algorithm

Vector Semantics: Word Meaning, Word Vectors, Co-occurrence matrices, Term-Document matrix, Cosine Similarity for words and documents, TF-IDF Model, Sparse and Dense Word Vectors, Word2Vec with Skip Gram Algorithm for generating word embeddings, properties of word embeddings

Neural Networks: Basic Neural Unit, Neural Network Architecture with hidden layers, activation functions, weights and biases, Back Propagation Algorithm for Gradient Descent in Neural Networks, use of word embeddings in neural networks

Recurrent Neural Networks: RNN architecture, RNNs for processing text documents, Building Language models using RNNs, Applications of Language Models, Back propagation for RNNs

Vanishing and exploding gradients, gradient clipping, LSTMs, Bidirectional LSTMs, GRUs

Overview of Statistical Machine Translation, Neural Machine Translation, Seq2Seq Architecture, Attention

Bias and Variance, Reducing bias and variance in neural networks, Dropout.

CNNs, Bert models.