LING/CSC 439/539: MIDTERM REVIEW OUTLINE

MIHAI SURDEANU

At the exam:

- You can bring a 2-page sheet with your notes. Note: if the exam includes the computation of derivatives, the instructor will provide all the differentiation rules necessary.
- You are allowed a simple, self-contained hand-calculator. Internet-connected devices are **not allowed** under any circumstances.
- Topics **in blue** are required only if covered during the 10/10 lecture.

Topics to know for the midterm:

1. Lecture 2: ML for Text Categorization
   A. What is text categorization?
   B. Applications of text categorization
   C. From data to vectors; feature generation
   D. Vector length, vector normalization, vector similarity (cosine similarity)
   E. K nearest neighbors (kNN)
      (i) Probabilistic kNN
      (ii) Cosine similarity and Euclidian distance
      (iii) Advantages, disadvantages
   F. Rocchio; why doesn't Rocchio generalize well?
   G. Perceptron
      (i) What is error-driven learning?
      (ii) Decision function
      (iii) Training algorithm
      (iv) Visualize what the Perceptron is doing during training

Date: Fall 2017.

(v) What do the values of the weights mean?

(vi) Simplifying the bias term

(vii) Voted Perceptron; implementation; advantages and disadvantages

(viii) Averaged Perceptron; implementation; advantages and disadvantages

(ix) Advantages and disadvantages of vanilla Perceptron

H. Logistic regression (LR)

(i) Decision function; sigmoid function, its derivative

(ii) Cost function for LR; how to obtain the gradient vector

(iii) Gradient descend/ascend algorithm; batch, stochastic

I. Feed-forward neural networks

(i) Architecture of feed-forward networks; w, b, a, z, δ

(ii) MSE cost function

(iii) BP1, BP2, BP3, BP4: formulas, meaning

(iv) Proofs of BP1 – 4

2. Lecture 2b: Best Practices

A. How to implement multi-class classification for: Perceptron, LR, and NNs

B. Regularization: L2, L1, dropout

C. Other decision functions: tanh, hard tanh, ReLU

D. Other cost functions: binary hinge, multi-class hinge, cross-entropy

E. Why is cross-entropy better than MSE?

F. Mini-batch SGD

G. Intuition behind momentum optimizers

H. Intuition behind adaptive learning rate optimizers

I. Xavier weight initialization

J. Vanishing and exploding gradients

K. Saturated and dead neurons

L. Learning rate scheduling

3. Lecture 3: Distributional Similarity and Word Embeddings

A. Distributional hypothesis; how to build it; advantages and disadvantages

B. Co-occurrence matrix

C. Latent semantic indexing

(i) Formula; what does each matrix mean?

(ii) Advantages and disadvantages

D. Word2vec, skip-gram

(i) Objective function; implementation using "center" and "outside" vectors

(ii) How to add negative examples to the objective function

(iii) Intuition behind continuous bag of words (CBOW) model; why is it worse?

4. Lecture 4: Linguistic Essentials

A. Tokenization issues, for English and other languages

B. How to design a sentence segmentation algorithm

C. Syntactic parsing

(i) Constituent trees; structural ambiguity

(ii) Dependency syntax; advantages and disadvantages

(iii) Generating dependencies (without head-word identification rules)

(iv) Important Stanford dependencies; structural ambiguity

(v) Shallow syntax intuition and example

D. Part-of-speech (POS) tagging

(i) English word classes

(ii) Penn Treebank tagset

(iii) Ambiguities in POS tagging

E. Word normalization

(i) Explicit vs. implicit classes

(ii) Porter's stemming algorithm; implementation intuition; advantages and disadvantages

(iii) Lemmatization; advantages and disadvantages

F. Named entity recognition (NER)