CSE 325/425 (Spring 2021) Homework 3

Due on 11:55pm, Mar 17, 2021

Grading: All questions have the same points (25 each). We will randomly grade some of the questions.

Submitting: Only electronic submissions on Coursesite are accepted. You can handwrite your answers on papers and then scan them to images. If you need to plot figures using a computer, the plotted files should be saved and included in the submitted pdf file. Submit a single pdf file named

<Your LIN>HW3.pdf

Other format will not be accepted.

Questions:

1. List the probabilities that HMM, MEMM and CRF need to estimated during model training, respectively. Don't just write down symbols or equations from the slides, but explain how to sum these probabilities to the constant 1 (in other words, point out the sample space of each probability distribution).

2. Design one feature function of the multi-class logistic regression POS-tagger, so that person's names can be tagged as NNP accurately.

   (*Hints: refer to the slides of lecture 9 and consider what characteristics of persons' names in a sentence will distinguish them from other words.*)

3. Based on the feature vectors and parameter vectors and the training sentence in Figure 1, take one gradient descent step to update the parameter vector $\boldsymbol{\theta}_{\text{NN}}$ at the word $o_t = \textit{Race}$.

4. For an RNN with

$$\mathbf{a}^{(t)} = \mathbf{b} + W\mathbf{h}^{(t-1)} + U\mathbf{x}^{(t)} \quad (1)$$
$$\mathbf{h}^{(t)} = \tanh(\mathbf{a}^{(t)}) \quad (2)$$
$$\mathbf{o}^{(t)} = \mathbf{c} + V\mathbf{h}^{(t)} \quad (3)$$
$$\hat{\mathbf{y}}^{(t)} = \text{softmax}(\mathbf{o}^{(t)}) \quad (4)$$

   Assume that $\mathbf{o}^{(t)}$ is of length 2, $\mathbf{h}^{(t)}$ is of length 3, and $\mathbf{x}^{(t)}$ is of length 2. Write down the dimensionalities of the parameters $\mathbf{b}$, W, U, $\mathbf{c}$, and V.

Secretariat	is	expected to		race	tomorrow.
NNP	VBZ	VBN	TO	VB	RB

		f_1	f_2	f_3	f_4	f_5	f_6
$c = \text{NN}$	$\mathbf{f}(o_t, \text{NN})$	1	0	0	0	0	1
	$\boldsymbol{\theta}_{\text{NN}}$	0.8	1	-2	3	0.1	-1.3
$c = \text{VB}$	$\mathbf{f}(o_t, \text{VB})$	0	1	0	1	1	0
	$\boldsymbol{\theta}_{\text{VB}}$	0.9	0.8	-1	0.01	0.1	0
$c = \text{VBG}$	$\mathbf{f}(o_t, \text{VBG})$	0	0	0	0	0	0
	$\boldsymbol{\theta}_{\text{VBG}}$	1	0.3	9	0.3	-0.4	-3.4

Figure 1: Features and parameters

5. Based on your answer to the above question, populate the parameters with all 1's (that is, all parameters are matrices/vectors of all 1's). Then execute the above equations for the two steps on two input vectors $\mathbf{x}^{(1)} = [1, 1]^\top$ and $\mathbf{x}^{(2)} = [2, 2]^\top$.