Assignment 3

(Due: Nov 30, 6am)

Question 1: Let $\mathbf{Q} \in \mathbb{R}^{N \times d}$ denote a set of N query vectors, which attend to M key and value vectors, denoted by matrices $\mathbf{K} \in \mathbb{R}^{M \times d}$ and $\mathbf{V} \in \mathbb{R}^{M \times c}$ respectively. For a query vector at position n, the softmax attention function computes the following quantity:

$$\mathrm{Attn}\left(\mathbf{q}_n, \mathbf{K}, \mathbf{V}\right) = \sum_{m=1}^{M} \frac{\exp\left(\mathbf{q}_n^\top \mathbf{k}_m\right)}{\sum_{m'=1}^{M} \exp\left(\mathbf{q}_n^\top \mathbf{k}_{m'}\right)} \mathbf{v}_m^\top \coloneqq \mathbf{V}^\top \mathrm{softmax}(\mathbf{K}\mathbf{q}_n), \tag{1}$$

which is an average of the set of value vectors $\mathbf{V}$ weighted by normalized similarity between different queries and keys.

Question 1a: Please briefly explain what is the time and space complexity for the attention computation from query $\mathbf{Q}$ to $\mathbf{K}, \mathbf{V}$, using the big O notation.

Question 2: Consider a probabilistic context-free grammar with the following rules (assume that S is the start symbol):

S → NP VP	1.0
VP → Vt NP	0.7
VP → VP PP	0.3
NP → DT NN	0.8
NP → NP PP	0.2
PP → IN NP	1.0
Vi → sleeps	1.0
Vt → saw	1.0
NN → man	0.1
NN → woman	0.1
NN → telescope	0.3
NN → dog	0.5
DT → the	1.0
IN → with	0.6
IN → in	0.4

Question 2a: What's the most likely parse tree for the following sentence under this PCFG? Show CYK chart you developed below.

the man saw the woman with the dog

Question 2b: What's the (marginal) probability of the following sentence under this PCFG?

the man saw the woman with the dog

Question 3: A trigram language model is also often refered as a second-order Markov language model. It has the following form:

$$P(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i \mid X_{i-2} = x_{i-2}, X_{i-1} = x_{i-1})$$

Question 3a: Could you briefly explain the advantages and disadvantages of a high-order Markov language model when comparing with the second-order one?

Question 3b: Could you give some examples in English where English grammar suggests that the second-order Markov assumption is clearly violated.