BBM 495
Introduction to Natural Language Processing
Midterm 2 – 21.05.2020 (Duration: 180 min)

Name-Surname:
Student no:

I attest that I have not given or received aid in this examination.

Signature:

Part I: Short Answers (each 4 points):

a) State one drawback of using chain rule. What is your solution to tackle with this drawback in language models?

b) State one inadequacy of language models that we discussed in the class. Give an example in English. What is your solution to tackle with this inadequacy?

c) We already learned that every regular expression can be converted into a finite state automaton. Prove it with an example along with a conversion for a language with two letters in the alphabet: {x, y}.

d) Give at least two methods that can be used for estimating the semantic similarity between words? Only the ones learned in the class will be graded. Compare their advantages and disadvantages.

e) Describe what "You shall know a word by the company it keeps" (Firth, 1957) means and give an example in English to support your description.

f) How is the conditional probability between two words is predicted by word2vec (Mikolov et al. 2013)? Describe mathematically by giving what each symbol refers to in your equation. Giving only the equation will not be graded.

Part II. Design (each 10 points)

a) Design a trigram PoS tagging model using Hidden Markov Models. How will it look like? Draw properly. How will we predict the best tag sequence in that tagger? Describe mathematically with your equations.

b) Draw a finite state transducer (FST) for the English rule that drops 'y' at the end of a verb when it takes a past tense suffix. For example, 'copy+ed' becomes 'copied', 'study+ed' becomes 'studied', 'cry+ed' becomes 'cried', etc. However, 'obey+ed' remains 'obeyed', 'play+ed' remains 'played' etc., because of the vowel before the 'y' letter. Exceptions will not be handled. Only neat and readable drawings will be graded.

c) You will build a text classification system. A number of articles on various topics are given. You will classify them into following categories: biology, geography, architecture, etc.

- What are the pre-processing tasks you will perform before training your model? Give your reasons. Are you going to use any tool/library for any of those tasks?

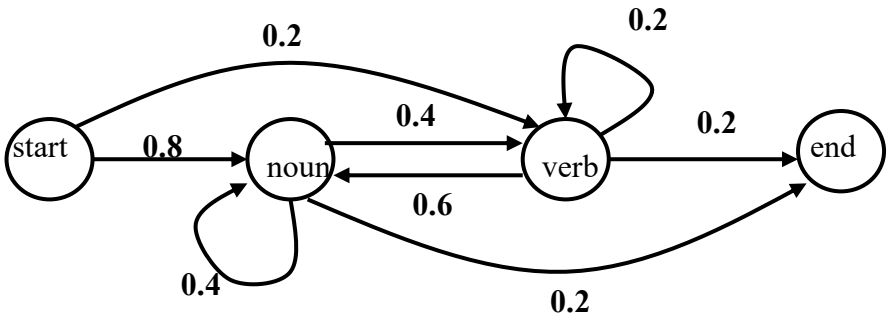
- Describe your machine learning model. You can define it either mathematically, or by drawing the architecture of your model along with a brief description of your equation or the architecture. Only the methods learned in the class will be graded.

Part III: Problem Solving

a) (8 pts) We have a two word language: "fish" and "saw". Suppose in our training corpus,

- "fish" appears 4 times as a noun and 2 times as a verb.
- "saw" appears 6 times as a noun and 8 times as a verb.

Decide which word belongs to which part-of-speech tag according to the given hidden Markov model by applying Viterbi algorithm. Show your work. Brute force solutions will not be accepted.



	start	**fish**	**saw**	**end**
Start	1			
Noun	0			
Verb	0			
End	0			

b) (9 pts) Give regular expressions for the languages generated by the following context-free grammars for a two-letter alphabet: {0, 1}:

a) S -> S1S1 | 0

b) S -> SSSSS | 1 | 0

c) (8 pts) Apply CYK algorithm for the following example with the given grammar:

CNF grammar G (with capital letters are non-terminals and lowercase letters are terminals):

- S -> AB | BC
- A -> BA | a
- B -> CC | b
- C -> AB | a

Is ababa in L(G), where L refers to a language with the specified grammar G?

d) (4 pts) According to the given feature vectors for the target words *dog*, *cat*, *car*, and *bark*, calculate the similarity between 'dog' and 'cat' by using cosine similarity. Are these two words semantically similar to each other according to your solution? Explain briefly. (You can do approximations in your calculation.)

	Leash	**Walk**	**Run**	**Owner**	**Leg**	**Bark**
Dog	2	4	2	8	2	2
Cat	0	2	4	2	2	0

e) (7 pts) You are a robot in an animal shelter, and must learn to discriminate Dogs from Cats. You choose to learn Naïve Bayes classifier. You are given the following examples:

Example	**Sound**	**Fur**	**Color**	**Class**
Example #1	Meow	Coarse	Brown	Dog
Example #2	Bark	Fine	Brown	Dog
Example #3	Bark	Coarse	Black	Dog
Example #4	Bark	Coarse	Black	Dog
Example #5	Meow	Fine	Brown	Cat
Example #6	Meow	Coarse	Black	Cat
Example #7	Bark	Fine	Black	Cat
Example #8	Meow	Fine	Brown	Cat

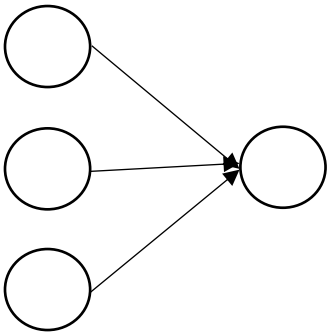
Consider a new example, (Sound=Bark, Fur=Coarse, Color=Brown). Is it a dog or a cat? Apply Naive Bayes algorithm.

f) (10 pts) A neural network with three input neurons, and 1 output neuron is given below. Suppose that the output neuron uses an identity activation function, i.e., $z = f(z)$ where z is the total input to the neuron. Let d be the desired output and d=4, and let $E = -d\ logz - (1 - d)\log(1 - z)$ be the cross entropy error. For the inputs (1,2,1) and weights (1,3,1), perform a forward pass to compute the output of the network and then apply backpropagation with gradient descent for one iteration. Write down the weights after the forward pass and also after the backward pass separately. Learning rate $\mu = 10$.

Hint: The derivative of the identity function is simply 1. The derivative of logarithm is given below:

$$\frac{\partial logu}{\partial u} = \frac{1}{u} \qquad \frac{\partial \log(1-u)}{\partial u} = \frac{-1}{1-u}$$

After forward pass:



After backward pass (backpropagation):