



Exam 2018, questions

Natural Language Processing (University of Sheffield)



Scan to open on Studocu



The
University
Of
Sheffield.

COM6513

Data Provided: None

DEPARTMENT OF COMPUTER SCIENCE

Spring 2018

NATURAL LANGUAGE PROCESSING

2 hours

Answer THREE questions.

All questions carry equal weight. Figures in square brackets indicate the percentage of available marks allocated to each part of a question.

1. a) What is a language model? Describe the data required train a language model, and what a trained model is expected to return.
[15%]
- b) What is the equation for the probability of a sentence? How is this probability approximated in an n-gram language model? Explain your terms.
[25%]
- c) What is add-1 smoothing? Why is it important for language modelling? Describe using equations how add-1 smoothing is applied to the bigram language model.
[20%]
- d) Describe back-off using equations to illustrate the key concepts. Explain how back-off is applied to language modelling making clear how we ensure that the resulting model is a valid probability distribution?
[20%]
- e) Language models can be evaluated intrinsically and extrinsically. Discuss the advantages and disadvantages for each approach and describe two methods for intrinsic and three for extrinsic evaluation.
[20%]

2. Information extraction is a well-known task in natural language processing. It is typically decomposed into two steps: named entity recognition and relation extraction. Consider the following sentences:

Michael Jordan, shooting guard for Chicago Bulls, was born in Brooklyn.
 Michael Jordan never played for his home city, Brooklyn.
 The hometown of Michael Jordan, CEO of PepsiCo, was Kansas City.
 Michael Jordan, native of Kansas City, studied in Yale.

- a) Using examples from the above sentences:
- (i) Describe how you could extract data to train a named entity recognition model. [10%]
 - (ii) Name two algorithms that would typically be used to learn named entity recognition models. [10%]
 - (iii) Describe the features that would typically be used by a named entity recognition system. [10%]
- b) Using examples from the sentences in the beginning of the question:
- (i) Describe how you could extract training data to train a relation extraction model for the relation `born_in`. [10%]
 - (ii) Name two algorithms that would typically be used to learn relation extraction models. [10%]
 - (iii) Describe the features that would typically be used by a relation extraction model. [10%]
- c) In order to populate a knowledge base containing the birthplaces of famous people we need to be able to link the named entities recognised with entries in the database, a task known as named entity linking.
- (i) Describe how you will extract training data to train a binary classifier to distinguish the mentions of Michael Jordan as a machine learning scientist from those of the CEO of PepsiCo. [10%]
 - (ii) Describe the features that would typically be used for this task. [10%]
- d) So far we have assumed separate models for each of the steps described, i.e. named entity recognition, relation extraction and named entity linking. Describe how the tasks could be solved jointly and what would be the benefits from doing so. [20%]

3. Consider the problem of topical text classification, i.e. deciding whether a piece of text talks about certain topic or not. A standard approach to the task is to learn a classifier using labelled data.

- a) Using equations, describe the model of logistic regression for classification. [20%]
- b) Using equations and pseudocode, describe how logistic regression can be trained using gradient descent. [30%]
- c) What features would be appropriate to represent the texts? [10%]
- d) Consider the following dataset:

class	text
sports	Wayne Rooney has announced his retirement from international football
sports	Joel Campbell, has returned from loan to play football
politics	This international committee promotes the elimination of racial discrimination
politics	13 years of UN negotiations aimed at restricting cyberwarfare collapsed

Given this dataset for training, the features you defined in question 3(c), and the logistic regression model combined with gradient descent to learn the weights, give 5 examples of features that should have weights indicative of the sports class and 5 examples of features that should have weights indicative of the politics class. Explaining why these features would be obtained from this dataset, and discuss whether they are useful for this task. Construct 2 examples where the model learned given this dataset and the features defined would predict the incorrect class; provide justifications for your examples.

[40%]

4. The perceptron and logistic regression are widely used classifiers in NLP.
- a) Compare the perceptron with logistic regression in terms of feature flexibility and score interpretation with reference to the equations used to define these classifiers. [20%]
 - b) Both the perceptron and logistic regression are linear models. Explain the limitation of linear models using suitable equations and diagrams. [20%]
 - c) Describe using equations and diagrams a multilayer perceptron with one hidden layer and explain how it can address the limitation of linear models. [30%]
 - d) Use pseudocode to describe the backpropagation algorithm for the multilayer perceptron. [20%]
 - e) Explain why neural networks are prone to overfitting and how this problem can be avoided. [10%]

END OF QUESTION PAPER