CSI4106 - Fall 2018
Preparation for final exam.

Exam type
Closed-book exam.
Calculator necessary. NO phone will be allowed.
Duration 180 minutes

Focus
The important content that you need to review is in the slides and what was presented in class in relation to these slides. Exercises done in class are also important. I will not ask questions on subjects that are only in the reading material.

Content before the midterm

For “what to know” about the first part of the semester (before the mid-term), see the document that was provided for midterm exam preparation. That is still valid. About 25% of the final exam will be on that first half of the semester. Since Propositional Logic was already a review for most of you, I am excluding it from the final exam. So focus on:

Part 2 - Searching in solution space (blind and heuristic searches)
Part 3 - Constraint Satisfaction Problem (consistency analysis + randomized searches)
Part 4 - Supervised Machine learning (decision trees / naive bayes)

Content after the midterm

I give below what you should know and the important slides. If you understand the material with the slides (and notes you took in class), you do not need to do the additional reading. The readings were meant to help you understand the material covered in class. I’ve covered in class everything that I will ask in the exam.

Part 5 - Intro to feed-forward neural networks (NN)

What you should know:

- Be able to describe the main components of a NN (input, output, weights, error function)
- Understand linear regression and logistic regression
- Know the formula for the sigmoid equation (I will not provide this equation)
- Be able to write the stochastic gradient descent (SGD) pseudo-code
- Provided I give you the formulas for the derivations (the ones we derived in class), be able to apply SGD for learning in linear and logistic regression
- Know the equations of the two types of errors (MSE and LogLoss) and know when it is likely to apply one equation or the other (I will not provide those equations)
- Provided the description of a situation (e.g. credit example, or like-travel example), be able to suggest basic architecture elements (input, output, loss function).
- Be able to describe what is input normalization, one-hot-encoding
- Understand what a Multi-Layer Perceptron is.
- Provided the derived formulas (as we did in class), be able to apply SGD for learning (changing weights) in a MLP (see exercise “to practice” in the slides CSI4106-NN-3)
- Understand what is overfitting, be able to describe a few factors causing overfitting
- Understand cross-validation

Important slides

CSI4106-NN-1
Intro: 6, 7, 8,
Perceptron: 14, 15, 16
Sigmoid function: 17
Linear regression: 20, 21,
Learning with gradient descent for linear regression: 23, 24, 27
Input normalization: 28
Logistic regression = Perceptron: 32
Learning with gradient descent for logistic regression: 33, 34 (error function), 35 (results of the derivation done in class), 36 (algorithm using these results)
Example of input/output and resulting model: 37, 38

CSI4106-NN-2
Input encoding: 10, 11
MLP architecture: 22, 23, 25
Non-linear activation functions: 26 (know the sigmoid + ReLU)
Learning with gradient descent in MLP: 28
Example of the difference between output of a logistic regression and a MLP: 31, 32

CSI4106-NN-3
Back-propagation: 6 (example in class) + 7, 8 (make sure to do this exercise)
Overfitting: 10, 11, 15
Cross-validation: 16, 17, 19

Part 6 - Natural Language Processing

What you should know:

- Be able to name a few tasks of interest for NLP
- Be able to name a few sources where text could come from for NLP analysis and the impact of these sources (e.g. an email could have more typographical error than a newspaper article)
- Be able to manually perform the steps in the NLP pipeline (the ones tested in the notebook): tokenization, stemming, lemmatization, POS tagging, sentence splitting
- For the POS tagging, no need to know all the tag set, but know the difference between Verb, Noun, Proper Noun, Preposition, Determinant, Adjective.
- Be able to describe and recognize problems of ambiguity for tokenization, sentence splitting, POS tagging and syntactic analysis.
- Understand what is Named Entity Recognition, and understand its three approaches: Regular Expressions, Gazetteers and supervised learning
- I will not ask to write regular expressions, but you should know what they would be good at detecting within the task of NER.
- In class, we designed two NN, one for Sentence splitting and one for NER. Be able to describe such NN (input/output) and provide various features.
- Understand what Edit Distance is and be able to calculate the edit distance between two provided strings. (I will not ask about Soundex)

Important slides

CSI4106-NLP-1
Tasks of interest in NLP: 8
Source of text for NLP: 17
NLP pipeline overview (highlighting ambiguity): 22, 23, 24, 25, 26
Tokenizer: 32
Stemming / Lemmatization: 33,34, 35, 36
POS tagging: 37
Sentence splitting: 38
Constituency parsing: 39, 40
*The notebook is also helpful for the NLP pipeline steps.

CSI4106-NLP-2
Ambiguity: 9, 10
Discussion on Inter-Annotator Agreement: 11
Building a NN model for sentence splitting: 21 (done in class)
Named Entity Recognition: 23, 24
Approaches to NER (RegEx, Gazetteers, Supervised ML): 28
Cases where regular expressions are good: 30
Examples of Gazetteers: 34, 35, 36
Edit distance: 37, 38
Output of a NER tagger: 40
Comparison of NER approaches: 42
+ Building a NN model for NER (done in class)

Part 7 - Knowledge Representation

What you should know:

- Understand what is Word Sense Disambiguation and Abbreviation Expansion
- Be able to describe what Wordnet is (talk of synsets and of its taxonomic structure)
- Understand the difference between a taxonomy and a semantic network
- Given a set of words, be able to construct a taxonomy (exercise done in class)
- Given a set of words, be able to construct a semantic network (exercise done in class)
- Understand the difference between a hypernym/hyponym relation and a part-whole relation
- Understand what a frame representation and be able to generate example sentences to illustrate what the participating elements in frames (exercise done in class)
- Understand the difference between surface form and concept
- Understand what a path-based similarity is in a semantic network
- Be able to perform the BOW-Overlap algorithm (presented in class and in notebook 7) and discuss the factors that will influence its results

Important slides

CSI4106-KR-1
Example of abbreviation expansion ambiguity: 8, 9, 10, 11
Example of common noun ambiguïty: 12
Wordnet : 16 to 22 (instead of the screenshots, best to explore online too, get a sense of its structure)
Taxonomy: 23
Additional relations in a semantic network, such as the part-of relation: 24
Semantic Network: 25
Representation of knowledge with predicates in DBpedia: 32, 33
Frames: 46, 48 (attention, error in Being_employed frame example, it's not Cook, it's Employer)

CSI4106-KR-2
Organizing knowledge example: 6 (exercise done in class)
Surface form variations: 8
Path-based similarity: 16, 17, 18
Definition-based similarity (Lesk): 20
Grounding (BOW-Overlap): 24, 25, 26 (BOW-Match is just a different name for BOW-Overlap)
Factors impacting the BOW-Overlap: 27, 28, 29, 30, 31, 32
* notebook for the BOW-Overlap

Part 8 - Deep learning architectures

What you should know:

- Be able to describe what an image auto-encoder does (dimensionality reduction, denoising)
- Understand what the difference is between a feed-forward NN and an auto-encoder
- Be able to describe what a GAN does and the main idea of a generator/discriminator working against each other
- Be able to describe the overall architecture of a convolutional network
- Be able to describe the 3 types of layers in a Convolutional Neural Network (convolution, pooling, fully-connected)
- Be able to apply a convolution of a filter on an image, followed by a max-pooling and show the resulting activation map
- Know what the stride is in the convolutional and max pooling layer, as well as the zero-padding

Important slides

CSI4106-DLA-1
Auto-encoders + denoising: 17-26
GANs: 42, 43, 44,
Example of result of GAN: 46, 47, 48

CSI4106-DLA-2
Input/output of a CNN for image processing: 11
Overall CNN architecture: 12
Convolution layer: 17, 18, 19
Multiple filters: 23
Convolutional + non-linearity: 20, 21, 24
Stacking Convolutional layers: 25
Stride and padding: 29-31
Pooling: 33
Overall architecture, including fully-connected layer: 36
Data augmentation: 46, 47