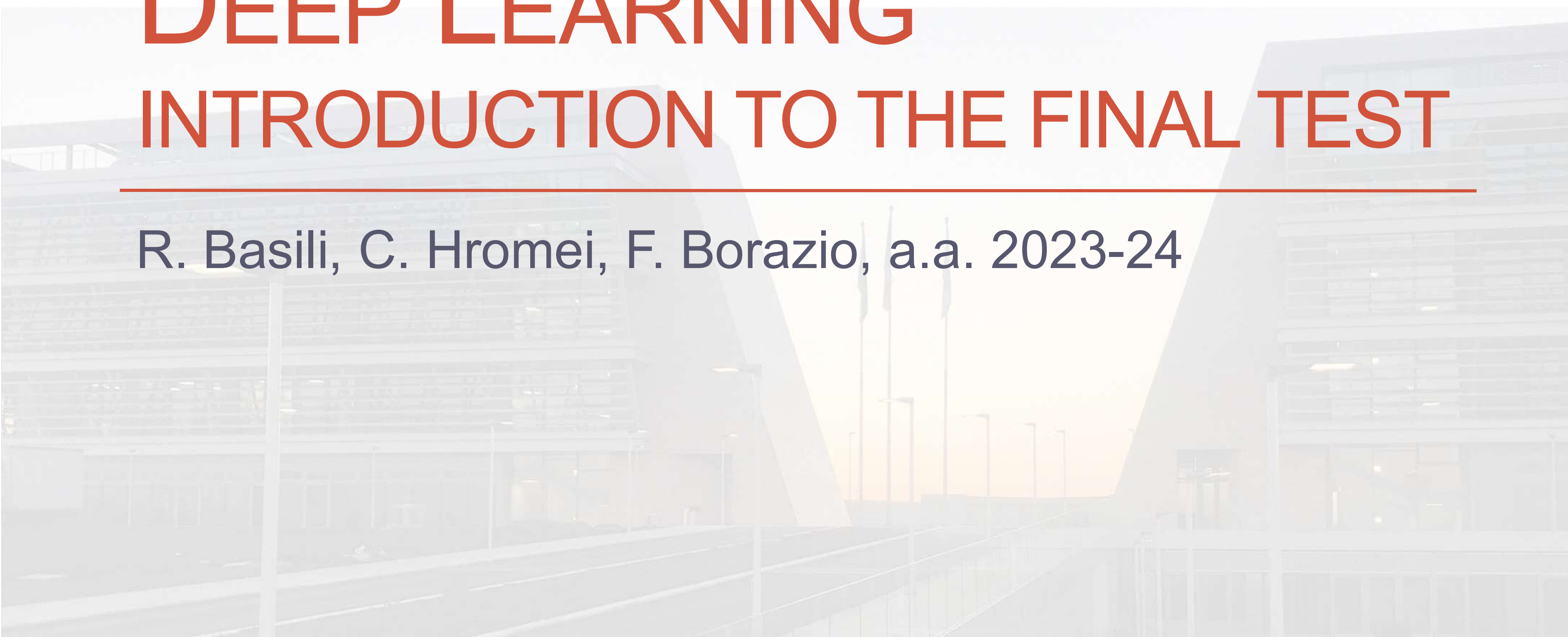
DEEP LEARNING

INTRODUCTION TO THE FINAL TEST



R. Basili, C. Hromei, F. Borazio, a.a. 2023-24

Overview

- Program Overview
- 2° Midterm:
  - Closed Question examples
  - Open Questions: examples
- Dates of the Exam
- Final Projects: impact on evaluation and modalities

Course Structure

- Three major topics
  - *Advanced topics in Machine Learning*
    - *HMMS,*
    - *Statistical Learning Theory, SVMs and Kernel Machines*
  - *Neural Networks and Deep Learning,*
    - *Convolutional Networks*
    - *Recurrent Neural Networks*
    - *Transformers*
    - *From encoder-decoder language models to Foundational Models*
    - *0-shot and few-shot Learning, Instruction Tuning and Prompting*
  - *Deep Learning applications: Document Classification, Natural Language Processing, Language Modeling, Question Classification, Named Entity Recognition, Textual Entailment, Question Answering, Sentiment Analysis*
- Different cross-relations between the three different sections:
  - Examples:
    - *Supervised Learning* (es. SVM) vs. Text Classification
    - HMM vs. *Language Modeling*
    - *Matrix decomposition* vs. *Mining of Lexical resources*
    - *Neural Learning for lexical embeddings*
    - *Convolutional Neural Networks for Image Processing*
    - *Encoding-Decoding for NLP tasks: from BERT to GPT-3*
    - *Semisupervised NN learning for NLP: pretrained language models*
    - *DL for Sentiment Analysis*

The program synthesis (1)

- **Lesson 0**: Deep Learning - a.a. 2023-24: Introduction: Course Organization and Exam Modalities.
  Short history of Large Language Models: perspectives for business processes.
- **Lesson 1**: Introduction to Web Mining & Retrieval.
  - Some slides of Lesson 1 refer to the discussion of the link: **"A visual introduction to ML"**, slide 24).
  - **Lesson 1.1: Machine Learning: target problems and major paradigms**.
- **Lesson 2**: Machine Learning Metrics and Evaluation (part I: metrics for Text Classification).
- **Lesson 3: Language Modeling - an Introduction to Hidden Markov Models for Sequence Labeling**.

Complementary Materials (Non mandatory):

  - **Lesson 3a**. Parameter Estimation for Language Modeling: the Baum-Welch algorithm.
  - **Lesson 3b**. Parameter Estimation and Rare Phenomena in Language Modeling.
- **Lesson 4**: (A gentle) Introduction to PAC learning and VC dimension.
  The slides used for the Course have been postedited from a kindly published version by Ethem Alpaydin, that you can find HERE.
  - **C. Burges's Tutorial on SVM and VC dimension.**
  - **D. Haussler discussion of PAC Learning**, 1999.
  - **Valiant L. G. *A Theory of the Learnable***, Communications of the ACM, Volume 27 Issue 11, Nov. 1984 Pages 1134-1142 .
- **Lesson 5 and 6**: Support Vector Machines and Kernels (Full package).
  - **An animated Perceptron**.
  - **Dan Klein's tutorial** on **Lagrange methods for the SVM optimization problem**.

The program synthesis (2)

- **Section II - Introduction to Neural Networks and Deep Learning Architectures**

Lesson 7 An Introduction to Neural Learning. The MultiLayer Perceptron: defining and training MLPs.

  - **Lab 1 - Introduction to Keras:** the XOR example.
  - **Lab 2a** - A Linear classifier and a MLP for image classification over the MNIST dataset in Keras.
  - **Lab 2b** - A Linear classifier and a MLP for image classification over the MNIST dataset in Pytorch.

- **Lesson 8 Deep Learning: Complex Architectures and Tasks: Convolutional Neural Networks**

References and resources:

  - Gradient Descent and Perceptron training, "*A Tutorial on Deep Learning, Part 1: Nonlinear Classifiers and The Backpropagation Algorithm*", Quoc V. Le (Google), (fino a sezione 6 compresa)
  - Backpropagation on NNs, "*Sparse autoencoder*", from the lessons of Andrew Ng (Stanford University, USA), (pages 1-12)
  - Advanced Architectures, "*A Tutorial on Deep Learning, Part 2: Autoencoders, Convolutional Neural Networks and Recurrent Neural Networks*", by Quoc V. Le (Google Brain)
  - Deep Learning, "*Learning Deep Architectures for AI*", Yoshua Bengio Dept. IRO, Universite' de Montreal, CA.
  - An interesting in-depth analysis on CNN
  - **Convolutional Neural Networks for Visual Recognition: CNNs for Visual Recognition on GitHub**

Software packages for Deep Learning:

  - TensorFlow
  - Anaconda
  - Torch
  - Keras
  - Tensorflow
  - Pytorch

- **Lesson 9: Recurrent Neural Networks, Encoder and Decoder Networks**
- **Lesson 10: Word embeddings as distributional semantic lexical representations.**
- **Lesson 11 Neural Word Embeddings: between Language Modeling and Lexical Acquisition**
- **Lesson 12 Attention: the introduction to Transformers**
- **Lesson 13 Deep Learning: NLP tasks, Benchmarking Datasets and Evaluation**
- **Lesson 14 Beyond Transformers: Decoder only Architectures.**
  From zero-shot and few-shot learning to large language model prompting.
- **Lesson 15 Beyond Transformers: Instruction Learning.**
  From zero-shot to Instruction Learning.
- **Lesson 16 LLMs Trends: LoRA and RAG.**

The program synthesis (3)

- Introduction to Sentiment Analysis

Laboratory Material and Exercises

- **Introduction to NLP with Spacy**: **Information Extraction from Texts: morphological information, normalized tokens (Lemmas), grammatical dependencies.**
  - Slides of the March 14th lesson
  - Python Notebook with exercises: running Spacy and questions.
  - Data set for Tests: Question Classification corpus.
- **Python Book for the exercise on HMM for POS tagging.** Development of the HMM model by probability estimation over a corpus. Development of the tagged corpus with Spacy.
- **Kernel-based Learning in KeLP.**
  Python Book for the exercise on HMM for POS tagging. Development of the HMM model by probability estimation over a corpus. Development of the tagged corpus with Spacy.
- **Application of CNNs to the MNIST dataset**: **Introduction to CNNs, overview and computation of Convolutional and Pooling layers with an application of CNNs on MNIST dataset.**
- **BIO tagging with LSTMs**: **The task of BIO tagging and Entity Extraction, Training a Long-Short-Term-Memory Network in Keras with Exercises**
- NEW! **Application of BERT model on Question Classification dataset**: **Overview of BERT model family and intro to Huggingface repository. Fine-Tuning a Linear Layer atop a Pre-trained BERT Model on the QC dataset, using the following dataset:**
  - **Question Classification dataset (both train and test set)**
- NEW! LangChain
  The **LangChain Framework for Sequence Tasks**: **Overview and Usage of the LangChain framework for sequences. Focus on the core modules: Models, Prompts and Chains.**
- NEW! 
  The **Application of Llama 3 model for relation extraction in the medical domain**: Python Notebook with the overview of the Relation Extraction task and management of the Huggingface repository. Exploit the Llama 3 capabilities investigating 0-Shot, One-Shot and Few-Shot paradigms, using the **Clinkart dataset**.

- ExtremeITA

Final Test on June 18°, h. 15:00

- Second Mid Term (on the second half of the program, Lessons 7-17)
  - Written test (on-line):
    - CAQ. 10/12 closed answer questions (about 30 mins)
    - OQ. 1 open question (about 90 mins)
    - Oral discussion (non mandatory): on OQ (+ CAQ error analysis)
- First Final Written test (on the full program) :
    - CAQ. 13/15 closed answer questions (about 40-45 mins)
    - OQ. 1 open question (about 90 mins)
    - Oral discussion (non mandatory) on OQ (+ CAQ error analysis)
- Final Oral discussion (due for 9 CFU, optional for the others):
  - The final project (max 2/3 people)

Topic of the second half

- 10/12 questions in the TRM
- Targeted Topics:
  - Neural Networks
    - Recurrent Neural Networks
    - Trasformers
    - Encoder-Decoder Architectures
    - 0-shot learning and Prompt Engineering
  - *Distributional semantics and Neural LMs*
  - *DL tasks: from Question Classification to Textual Entailment and Relation Extraction*
  - Opinion Mining

Overview

- Program Overview
- 2° Midterm:
  - Closed Question examples
  - Open Questions: examples
- Dates of the Exam
- Final Projects: impact on evaluation and modalities

Closed Answer Questions: examples

Questions about NNs

- What is the definition for a *convolutional neural network* and which are the main differences with *recurrent neural networks*
- What is a *non linear activation function*? What is its contribution to the training process?
- Which ones among the following techniques is specifically devoted to improve the model optimization stage of a NN (for example, by avoiding *overfitting*)?
  a) Dropout
  b) Early Stop
  c) Input normalization
  d) The Stochastic Gradient Descent

Questions about NNs

- Discuss the notion of *Loss Function* that characterizes the training of a neural network
- The backpropagation method allows to:
  a) Improve the efficiency of the learning process of a NN
  b) Maximize the Loss function over the validation data set
  c) Minimize the Loss function over the training data
  d) Maximize the Loss function over the training data

- What is a *Recurrent Neural Network*?
- How can we control the *number of epochs* required to train a NN?

Questions (cont)

- What is the advantage in adopting the *mini-batch* policy during the training stage with respect to batches based on one single instance?
- Formalize the *Stochastic Gradient Descent* algorithm and discuss its objectives.
- The *Back Propagation through Time* technique allows to maximize the Loss function in:
  a) Recurrent Neural Networks
  b) Convolutional Neural Networks
  c) Neural Networks made by a single perceptron
  d) None of the Other
- What is the *Vanishing Gradient*?

Latent Semantic Analysis (LSA) (1)

- Let M= ((1 -1) (1 1) (-1 1)) be the initial co-occurrence matrix (Vocabulary V={t1,t2}). Determine the value σ_1 of the largest singular value
- R1: It is not possible: the problem is *under determined*
- R2. $\sigma_1 = 2$
- R3. $\sigma_1 = 1$
- R4. $\sigma_1 = \sqrt{2}$

Closed Questions (3)

- Determine the correct definition for the sentiment classification task amon the following:
  - (A) At document level this task corresponds to sentence classification into positive, neutral and negative polarity classes
  - (B) At sentence level the task consists into recognizing the features of objects to which the sentence sentiment makes reference
  - (C) At sentence level there are two tasks: (1) identification of subjective sentences in the input text; (2) polarity classification of individual sentences
  - (D) The task consists in the grouping of synomim expressions by which opinion holders may make reference to the object features
  - (E) None of the others corresponds to an acceptable definition

Closed Questions (4)

Signal the correct answer among the following ones:

a) Sentiment Analysis over Twitter is generally a simple task as the text corresponding to a *tweet* is limited in size.

b) User opinions in the social networks are not much interesting for the companies.

c) Sentiment Analysis is the computational study of opinions and sentiment expressed in texts.

d) Sentiment Analysis only rely on *machine learning* algorithms.

e) Sentiment Analysis is the computational study of the opinions and sentiment espressed by the topic of a text (e.g. an event or an entity)

Closed Questions (5)

Determine the correct definition among the following:

a) *Distributional* semantics methdos (e.g. LSA or *wordspaces*) cannot be adopted for the *relevance feedback* methods as they use vectors as representation models for terms.
b) *Distributional* semantics methdos cannot be adopted for the *relevance feedback* methods as they use lexical objects (i.e. symbols in the word dictionary) as representation models for terms and cannot be combined algebraically
c) None of the others
d) With *relevance feedback* we can impact only performances in terms of improvements in *precision*.

Questions about CNNs

- In a convolutional layer composed by 10 filters each of a dimension 3x3, and a stride value equal to 1, what is the number of parameters to be fine-tuned during training :

- 1) 100
- 2) 90
- 3) 9
- 4) 900
- 5) None of the other

Other Questions about CNNs

- What is the tensor dimension of the output of a convolutional layer made of 10 filters whose type is 3x3 to an image of dimension 28x28 with a stride equal to 1 and a null padding (i.e., 0 padded cells).

- 1) 6760 (26x26x10)
- 2) 676
- 3) 7840 (28x28x10)
- 4) 10
- 5) None of the others

Closed answer Questions on Labs

- Consider a sentiment classification task where the possible sentiments are *Positive*, *Neutral*, and *Negative*. Given the following code snippet:

```
model_name = "distilbert-base-cased"

class Classifier(nn.Module):
    def __init__(self, model_name, num_labels=2, dropout_rate=0.1):
      super(Classifier, self).__init__()

      self.encoder = AutoModel.from_pretrained(model_name)

      config = AutoConfig.from_pretrained(model_name)
      self.cls_size = int(config.hidden_size)

      self.input_dropout = nn.Dropout(p=dropout_rate)

      self.fully_connected_layer = nn.Linear(linear_layer_size,num_labels)
```

choose the correct answer.

1. It is not possible apply `Dropout()` in a Transformer encoder model like BERT
2. The code is incorrect because, for this classification task, it is more effective to use a linear layer on top of each embedded wordpiece.
3. `linear_layer_size = 3, num_labels = self.cls_size`
4. `linear_layer_size = self.cls_size, num_labels = 3`
5. `linear_layer_size = 3, num_labels = 3`

Closed answer Questions: solutions

Questions about NNs

- What is the definition for a *convolutional neural network* and which are the main differences with *recurrent neural networks*
- What is a *non linear activation function*? What is its contribution to the training process?
- Which ones among the following techniques is specifically devoted to improve the model optimization stage of a NN (for example, by avoiding *overfitting*)?
  a) Dropout
  b) Early Stop
  c) Input normalization
  d) The Stochastic Gradient Descent

Questions about NNs

- Discuss the notion of *Loss Function* that characterizes the training of a neural network
- The backpropagation method allows to:
  a) Improve the efficiency of the learning process of a NN
  b) Maximize the Loss function over the validation data set
  c) Minimize the Loss function over the training data
  d) Maximize the Loss function over the training data

Questions (cont)

- What is a *Recurrent Neural Network*?
- How can we control the *number of epochs* required to train a NN?
- What is the advantage in adopting the *mini-batch* policy during the training stage with respect to batches based on one single instance?
- Formalize the *Stochastic Gradient Descent* algorithm and discuss its objectives.
- The *Back Propagation through Time* technique allows to maximize the Loss function in:
    a) Recurrent Neural Networks
    b) Convolutional Neural Networks
    c) Neural Networks made by a single perceptron
    d) None of the Other
- What is the *Vanishing Gradient*?

Questions (cont)

- The *Back Propagation through Time* technique allows to maximize the Loss function in:
    a) Recurrent Neural Networks
    b) Convolutional Neural Networks
    c) Neural Networks made by a single perceptron
    d) None of the Other
- What is the *Vanishing Gradient*?

Answers

Overview

- Program Overview
- 2° Midterm:
  - Closed Question examples
  - Open Questions: examples
- Dates of the Exam
- Final Projects: impact on evaluation and modalities

Exam: Open Question Topics

- Program section concerning MidTerm 1.
  - Generative Models
    - Modeling Sequence Labeling Tasks through generative models
    - Estimating probabilities for SLTs
  - Applications of Automatic Classification: a comparative discussion
  - Statistical Learning Theory
    - Support Vector Machines
    - Kernels
  - Latent Semantic Analysis

Exam: Open Topics (2)

- Program section conerning MidTerm 2
  - Neural Network Learning
    - MLP, CNNs, RNNs,
    - Transformers
    - Encoder-Decoder architectures: Motivations and Objectives.
    - Examples of Encoders (e.g. BERT) and decoders (GPT)
    - Applications
  - Program Section 2: Statistical NLP and Social Web applications.
    - Embedding for Web Applications
      - Wordspace and their applications to search engines or NLP.
      - Motivations and different techniques for word embeddings
    - Statistical NLP and its applications in the Web
      - Semantic role labeling (as a sequence classification task)
      - Sentiment analysis Opinion Mining
      - Question Answering

Open Questions: examples

- Targets ML methods:
  1. Kernel methods
  2. Methods of *neural learning*
  3. Word spaces (for Query Expansion) and Word Embedding through NNs
  4. Sentiment Analysis on movie reviews

- Request:
  - Define basic methodological assumptions of the problem (model assumptions, type of observations available, external resources)
  - Describe the pseudo-algorithm or the functional architecture adopted for solving the task
  - Discuss the possible evaluation metrics
  - Discuss the potential applications of the proposed solution

Open Questions

1. Discuss the main differences between Support Vector Machines and the approach of Neural Networks in supervised learning tasks
2. Discuss the main architectures for deep learning and their differences, describing also their main applications
3. Please define and discuss the notion of kernel functions in the area of Statistical Learning Theory. Provide examples of their application to classification problems in the Web or NLP tasks
4. Please determine the main computational challanges related to the task of Sentiment Analysis and discuss potential resources used to approach them.

Final Test: Example of Open Question

Please discuss the application of ML methods (such as markov models or NNs) to *sequence labeling tasks*. (Make use of an example through an applications, such as POS tagging of natural languae sentences)

- Request:
  - Define basic methodological assumptions of the approaches
  - Define the basic notions of hidden states, transitions and emission
  - Define the general model equations and discuss the algorithmics
  - Discuss the possible evaluation metrics

Variants

- Apply the HMM or NN modeling to the problem of URL recognition in free texts.
- Please make use of the state lables such as IOB for the start (B), inner (I) and outern (O) elements of a valid URL.
- Define the state vocabulary, the transition and emission matrices. Discuss possible parameter estimate techniques and their corresponding challenges.

Open Questions: comments on paper

- Read and discuss the following paper
  - Convolutional Neural Networks for Sentence Classification, Yoon Kim, Proc. of EMNLP 2014.
  - Positional Attention-based Frame Identification With Bert: A Deep Learning Approach To Target Disambiguation And Semantic Frame Selection, 2019
  - EMPATHBERT: A BERT-based Framework for Demographic-aware Empathy Prediction, 2021
  - Wenpeng Yin, Jamaal Hay, and Dan Roth. 2019. Benchmarking Zero-shot Text Classification: Datasets, Evaluation and Entailment Approach. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 3914–3923, Hong Kong, China. Association for Computational Linguistics.

  - Targets:
    - The task (e.g. TC, NERC, RTE, SA, …)
    - The datasets,
    - The NN model, the resources (e.g. pretrained BERT models),
    - Main results, Observations/criticisms of the paper
    - Link the Course Program sections used to understand the main paper contents

Overview

- Program Overview
- 2° Midterm:
  - Closed Question examples
  - Open Questions: examples
- Dates of the Exam
- Final Projects: impact on evaluation and modalities

DL a.a. 2023-24: Summer Exam Session

- Second Mid Term (*):
  - **June 18°, 2024**, **Aula *tbd*, h. 15:00-17:30**
  - **Oral discussion: around June 22nd, 2023**
- First Final test (*):
  - **June 18°, 2024**, **Aula *tbd*, h. 15:00-17:30**
  - **Oral discussion: around June 22nd, 2023**
- Second Final test (*):
  - **15 July 2023**, **h. 10:00-13:30**
  - **Oral discussion: around July 20, 2020**

- Lab exercises: optional before the final (or MidTerm2) written tests
- Projects (mandatory for 9 CFU):
  - Project presentation (after the Final Test)
  - Discussion via slides (ca 20 mins), when finished, whenever agreed upon with the teachers (via mail)

Overview

- Program Overview
- 2° Midterm:
  - Closed Question examples
  - Open Questions: examples
- Dates of the Exam
- Final Projects: impact on evaluation and modalities

Final Projects:

- Prompt Engineering for Relation Extraction
- LowRank approximation: performance analysis
- Medical IE from noisy texts
- Multimodal systems: Large Vision Language Models
- Explainable AI decision through Large Language Models:
  - Combining Prompt Engineering and Textual Entailment
  - Adoption of Graph Neural Networks in combination with LLMs