

# CSA5024 - Advanced Methods in LCT:

## Guest lectures in NLP & deep learning

Malta, March 2019

Barbara Plank

# Agenda

- Course logistics
- Introduction to FFNN
- Project topics

# Course logistics

- Lecturer: Barbara Plank (bplank@gmail.com)
- Time:
  - Monday 15-18
  - Tuesday 11-13
  - Thursday 11-13
- Location: Block A, room 17 (tomorrow: 16)

Introduction round

# Course outcomes

- After completing this course, students are able to tackle a challenge in NLP, implement a solution, and critically assess their solution in the light of recent research papers in NLP
- Focus: Deep Learning for NLP

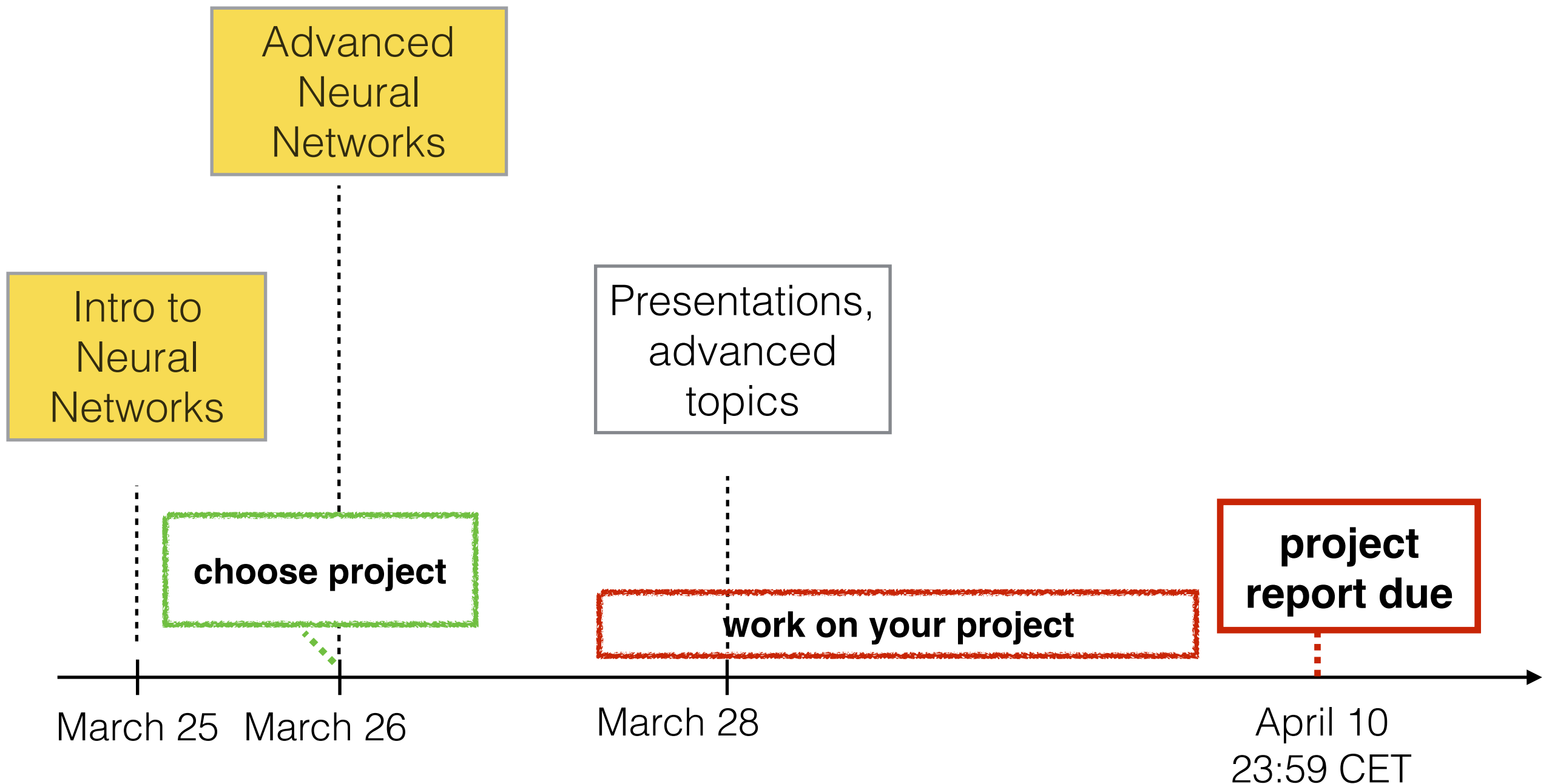
# Prerequisites

- Familiarity with Machine Learning and Natural Language Processing
- Knowledge of Python - there is a tutorial [here](#)

# Grading

- 100% project report

# Course timeline





# Tools

- python3 (suggestion: Anaconda)
- Sk-learn (today, I assume you are familiar with it)
- Keras and Tensorflow  
(you can use others if you are more familiar with them, e.g., DyNet)

# Course page

<https://github.com/bplank/2019-ma-notebooks>

# Project topics

# General Theme

- **Comparison** of a traditional method (e.g. LogRegression, SVM, HMM/CRF...) to a neural network approach
  - Baseline: traditional ML approach
  - Your system: neural network
- You can work in groups (2 people)

# Project Topics

- Gender prediction on Twitter
  - Language independent
  - Cross-lingual
- Swiss Dialect Identification
  - Including acoustic data
- Amazon reviews sentiment classification over different domains

# Language-independent Gender Prediction on Twitter

A user has posted the following tweets:

- USER sterkte en snel veel beterschap !
- Dio Oberon overleden: 'de beste nar van Nederland' - Show - Algemeen - bndestem URL
- USER dat is jullie pap op tv! Leuk!
- ja hoor dit gaan we winnen #songfestival ! ;-(((
- USER haha precies wat ik ook dacht ! Watje :-)
- USER hoe moet dat nou een feestje in Rooi zonder jou? beterschap USER xxx ook van katjazondertwitter :-)
- op de bank (alleen) :-)
- USER hè toch! Beterschap!
- Ik heb op deze video gestemd bij #Dichtbijfonds, jij ook? URL
- USER mooi ! ik wil ook nog deze week maar weet niet of ik de tijd kan vinden ;-( vrijdag of zat. hoop ik !!!
- MooiRooi.nl - Sint-Oedenrode - U kunt zich nog aanmelden als wensvervuller URL
- Shit gestoken door een #wesp au!
- USER vermaak je je een beetje?
- Kliknieuws.nl | Sint-Oedenrode - Foto's Hans Vervloed in bibliotheek URL
- USER USER hoe is het met jullie? nog gegaan en drank gewonnen? Ik kon dus echt niet komen #pijn ! morgen over hoop ik :-)
- USER USER was het een leuk feestje ? nog een fijne avond x
- Beleef een professionele fotoshoot bij Lachebekkies. Inclusief foto's op verschillend forma.. URL Kei leuk !!!!!!!
- USER USER USER USER weer naar huis :-) bedankt allemaal voor de gezelligheid xxx URL
- baaah dan blijf je echt liever alleen !!
- Film tip ! Pina #Pinabausch in 3D #modernballet. heel indrukwekkend !

**Do you think that the poster of these tweets is male or female? (required)**

- ☐ Female
- ☐ Male

# Gender Prediction on Twitter

- Data: Subset of Twisty corpus (prepared for you) covering EN, NL, FR, ES, PT
- Approach 1: Using language-independent features Ljubesic et al., 2017
- Approach 2: Cross-lingual gender detection (van der Goot et al., 2018 ACL)

# Swiss Dialect Identification

- Can we detect where a speaker is from (in Switzerland)?
- 4 dialect areas:
  - Bern (BE), Basel (BS), Lucerne (LU) and Zurich (ZH)
  - VarDial 2019 shared task (closed)
- Available data:
  - Transcriptions
  - i-vectors (400-dimensional iVector representing the acoustic features of the utterance)
  - Automatically normalized text



# Amazon Sentiment

- Large-scale review dataset
- How to build a robust sentiment analysis system?