

Machine Learning Practical

Part 2: Group Projects

MLP Part 2: Group Projects

Steve Renals

Machine Learning Practical — MLP Lecture 11
24 January 2018

<http://www.inf.ed.ac.uk/teaching/courses/mlp/>

- One introductory lecture (today)
- Questions and answers (next week - 31 January)
- Four guest lectures (weeks 4—7)
 - 7 February: Ben Allison, Amazon – Building Production Machine Learning Systems
 - 14 February: Hakan Bilen, University of Edinburgh – Unsupervised learning of object landmarks from equivariance
 - 28 February: Vincent Wan, Google – Speech synthesis using LSTM auto-encoders
 - 7 March: Subramanian Ramamoorthy, University of Edinburgh and FiveAI – Problems for Machine Learning Practitioners from the Autonomous Driving Domain
- Note: lectures on 7 Feb, 14 Feb, 7 Mar will not be recorded

Group projects

- Semester two will be based on **group projects**
- 2–3 students per group
 - You can discuss any aspects of the assignment with your group
 - Divide up the tasks any way you like
 - Best if the team collaborates on each part
 - If you haven't already, register your group at
https://docs.google.com/spreadsheets/d/1bS9kYr3E78Us8zdTt4SaVLZlBUAyn9_m83Ya3s7HiZs/edit?usp=sharing

Project scope

- We can give some pointers, but scope your own project:
 - Feasible to do in 7 weeks, in a group of 2–3, given you have other courses going on!
 - Needs to have a significant amount of experimentation
 - Should link to the main themes of MLP so far, but you can extend things
 - Conv nets, recurrent networks, feed-forward networks, ...
 - Classification, density estimation, reinforcement learning, ...
- How to choose a project?
 - Begin with an interesting data set or task, and focus on engineering fairly standard approaches to work well
 - Begin with a more challenging approach and work on a dataset you already understand and for which you have good baselines

Both types of project are valid, and you can get excellent marks on both types

Start by making a plan - what data you will be using, what approaches you will investigate, what are the research questions?

- Possible data sets
 - CIFAR-10/100 object recognition
 - Million Song Database (or a subset) for music genre recognition
 - Movie review dataset for sentiment analysis
 - Painter-by-numbers – predict if two paintings are by the same artist
 - Bring Your Own Data (BYOD)
- Possible approaches to explore
 - multitask learning
 - curriculum learning
 - one-shot learning
 - Bayesian deep learning
 - meta-learning
 - deep density estimation

Interactions with instructors

- Lectures until week 7
- MLP Helpdesk (weeks 2–9) – Monday-Friday, 14:00-15:00
AT 5.08 South Lab – best place for technical queries.
- Tutorials (weeks 3–9) – discuss the progress of your project
- Piazza – ask and answer questions, search for teammates, ...
- No scheduled labs this semester

- Each group is assigned to a tutor, who will discuss and review progress
- Set up a Google Doc for your group (shared with instructors)
 - report progress and experimental results, give plans, raise questions
- Weekly tutorial sessions to meet with tutor – tutorial sessions will involve 5-6 groups
- Update Google Doc at least 24 hours before tutorial session
- Will be a sign-up sheet for tutorials (soon)

Computing... “you’re gonna need a bigger boat”

Computing... “you’re gonna need a bigger boat”

- Deep learning uses up a lot of compute cycles...

Computing... “you’re gonna need a bigger boat”

- Deep learning uses up a lot of compute cycles...
- Introducing the MLP GPU system:
 - Available from start of next week
 - Initially with about 80 GPU (NVidia 1060 Ti) cards available
 - Next month there should be up to 200 GPU cards available (also 1060 Ti)
 - It’s a new system and we are all pre-alpha testers!
 - More details very soon

Computing... “you’re gonna need a bigger boat”

- Deep learning uses up a lot of compute cycles...
- Introducing the MLP GPU system:
 - Available from start of next week
 - Initially with about 80 GPU (NVidia 1060 Ti) cards available
 - Next month there should be up to 200 GPU cards available (also 1060 Ti)
 - It’s a new system and we are all pre-alpha testers!
 - More details very soon

Why 1060Ti? Need to make a balance between power consumption*, computer performance, and cost...

(*) When running 200 GPUs, the issue of power consumption becomes really important!

Coursework 3 – Interim Report

- Motivation and introduction to the project
- Aims and objectives – be precise
- Data set and task
- Research questions
- First phase of experiments
- Any interim conclusions
- Plan for the remainder of the project, including discussion of risks, backup plans
- Submission deadline: Thursday 15 February, 16:00

Coursework 4 – Final Report

- Brief introduction, including a reprise of the aims and objectives, the data and the task
- Experiments
 - Methodology
 - Results
 - Discussion and interpretation
- Conclusions and discussion
 - Conclusions with respect to aims and objectives, research questions
 - Any changes with respect to the original plans
 - Discussion of what was achieved and learned in the project
 - Potential further work
- Submission deadline: Friday 23 March, 16:00

- *Can I do the project alone?*

We won't stop you, but it is not recommended. We are expecting projects to be have the amount of work from a 2-3 person group; interacting with your team is an important experience.

- *Do we have to use TensorFlow?*

No: Keras, MXNet, PyTorch, ... would all be OK.

- *Can this be part of my dissertation project?*

No, it should be completely separate.

- *Can I use cloud services like AWS or Google Cloud?*

Yes, if you wish to

- **Weeks 1–2:** Introduction to TensorFlow – branch `mlp2017-8/mlp_tf_tutorial` on the MLP github
- **Week 2:** Form project groups
- **Weeks 2–3:** Scope projects, setup google doc, start work!
- **Week 5 (15 February):** Interim report
- **Week 9 (23 March):** Final report