# PHYS4035 MACHINE LEARNING IN SCIENCE PART 1 (2024)

## Project and Paper

You will work in groups of three or two (due to numbers and preferences) on a research project based on your preferences on selected aspects of ML. The outcome of the project will be a short paper describing the project and results in the style of a research journal article. The project allocations and descriptions are given below.

**Paper:** The project description and results will be presented in a **3 page paper** (2 pages for two-member group) in the style of Physical Review Letters, ideally typeset in LaTeX (for example via the RevTeX package) although we will accept other modes of preparation compatible with a two column PRL format and style. The paper will have the usual structure (title/authors/affiliations/abstract in the heading section, followed by the main text, figures with appropriate captions, standalone equations, and references). Do **not** include code in the paper (only pseudo-code if necessary to describe an algorithm).

**Groups doing projects that require coding:** All programming will be done in python (version 3). Use of neural network packages (Keras, TensorFlow, etc.) is **not** allowed; use of other standard packages (numpy, scipy, etc.) is allowed. The code will be submitted together with the paper. The readability of the code will form part of the assessment. (NB: both paper and code will go through plagiarism detection system.)

**Schedule and deadline:** Projects will be announced by Nov 28th (during the RL-W3 class). There are no online workshops after that week to allow time for project work. The deadline for handing in the paper is **February 3rd 2025 3pm (*new deadline*)** (submission via Moodle).

**Assessment:** The mark for the Project and Paper contributes 60% to the overall mark for the module. The assessment has two components

Quality and style of presentation = 50%
Content and understanding = 50%

**Project groups (An = CNeuro, Bn and Cn = MLiS):**

Group A1, Project P1: Johnson, Thomas-Roche, Undelikwo

Group A2, Project P2: Carr, Gunel, Srinivasan

Group A4, Project P4: Broadhurst, Kuntipalo, Su

Group A6, Project P6: Condon, Mathias, Oliver

Group A9, Project P9: Goldsmith, Hardy, Mahmoud

Group B1, Project P1: Pan, Xuan, Zhao

Group B2, Project P2: Barrick, Chongsawad, Patil

Group B4, Project P4: Jadhav, Li, Tam

Group B6, Project P6: Liu, Shen, Yan

Group B8, Project P8: Gamston, Rudd, Underdown

Group B9, Project P9: Atkinson, Marshall, Rolfe

Group C1, Project P1: Chandran, Kuatbekov, Lahane

**Project descriptions.** Here we provide a brief description of the projects. These are meant to set the overall topic and a basic specification of the problem to study. We expect students to research the topic and come up with a suitable project to pursue along the proposed lines.

### P1 Breast Cancer Tumours (UL)

The data for the Breast Tumours is available from the ML repository of UC Irvine

https://archive.ics.uci.edu/dataset/15/breast+cancer+wisconsin+original

under

*Breast Cancer Wisconsin (Original)*

The goal of the project is to build an understanding of the difference in features for malignant and benign tumours, using unsupervised learning methods, based on the features available in the above data set.

### P2 Butterfly Species Richness (UL)

Data for the Butterfly Species Richness is provided in the file **`ButterflyData.ods`** available in Moodle.

The goal is to understand the distribution of butterfly species richness from an USL perspective. In particular, understanding what features are important for determining the distribution. Using the location name, one can create a large number of features, and therefore a very high dimensional feature space. Unsupervised learning methods can be used reduce the dimensionality of this space, and therefore understand the correlations between these features and butterfly richness. Examples of techniques might include, PCA or autoencoders.

**P3 Artificial vs Neurological Learning (theoretical)**

Address the question: Is it necessary to re-learn networks with every task variant, or can we create a smarter network? Use the following reference as a starting point,

J.X. Wang, “Prefrontal cortex as a meta-reinforcement learning system”, Nature Neuroscience **21**, 860 (2018).

**P4 Classification of Breast Cancer Tumours (SL)**

The data for the Breast Tumours is available from the ML respository of UC Irvine

https://archive.ics.uci.edu/dataset/15/breast+cancer+wisconsin+original

under

*Breast Cancer Wisconsin (Original)*

The goal is to build a model to predict whether a given tumour is malignant or benign, based on the features available in the above data set.

**P5 Artificial vs Neurological Learning (theoretical)**

Address the question: How do deep networks for image recognition compare to the mammalian visual system? Use the following reference as a starting point,

D.L.K. Yamins et al., “Performance-optimized hierarchical models predict neural responses in higher visual cortex”, PNAS **111**, 8619 (2014).

**P6 Changes to Arctic Ice Extent (SL)**

The Arctic Ice Extent data is available at

https://noaadata.apps.nsidc.org/NOAA/G02135/north/monthly/data/

The goal is to build a model to predict the Arctic ice extent and, more ambitiously, the first time (if any) when the Arctic will be ice-free.

Potential Ideas and Approaches: Can start by simply building a linear regression model to predict for the year-averaged Arctic Ice extent with the year as input. Can continue development by finding more features to use as inputs, such as the global mean temperature or CO2 concentration, or using time-series approaches (predicting the next year’s extent based on previous extents). Can also try to model seasonality by including the monthly extent data, and predict the first ice-free year.

**P7 Reinforcement Learning in Neurological Systems (theoretical)**

Explore how much reinforcement learning in ML relates and is inspired by related problems in biological systems. A good starting point is this classic paper on reward signals in the brain.

W. Schultz, P. Dayan, and P.R. Montague, “A neural substrate of prediction and reward”, Science **275**, 1593-1599 (1997).

### P8 Reinforcement Learning of (stylised) Blackjack (RL)

The aim of this project is to train an agent to play the card game Blackjack, in an environment that can have a varying degree of complexity. For simplicity, we consider a stylised version of the game where there is no opponent. For details see description in Moodle.

### P9 Controlling a Drone via Reinforcement Learning (RL)

The aim of this project is to control an idealised drone that moves in two dimensions by means of reinforcement learning. For details see description in Moodle.