11th November 2025

PRACTICAL COURSE WORK FOR MACHINE LEARNING (H) COMPSCI 4061

Code of Assessment Rules for Coursework Submission

Deadlines for the submission of coursework which is to be formally assessed will be published in course documentation, and work which is submitted later than the deadline will be subject to penalty as set out below. The primary grade and secondary band awarded for coursework which is submitted after the published deadline will be calculated as follows:

1. In respect of work submitted not more than five working days after the deadline
   (a) the work will be assessed in the usual way;
   (b) the primary grade and secondary band so determined will then be reduced by two secondary bands for each working day (or part of a working day) the work was submitted late.
2. Work submitted more than five working days after the deadline will be awarded Grade H.

Penalties for late submission of coursework will not be imposed if good cause is established for the late submission. You should submit documents supporting good cause via MyCampus. Penalty for non-adherence to Submission Instructions is 2 bands You must complete an "Own Work" form via `https://studentltc.dcs.gla.ac.uk/` for all coursework

Submission and deadline

Your submission will be:

- a report in PDF which contains the answers to all questions including discussions, algorithms, plots and figures.
- the source code of all the implementations required to conduct the experiments.

Deadline: You must submit this on Moodle by 6.00 PM, Wednesday 3rd December 2025.

Applying Ensemble Learning for Breast Cancer Detection

As a part of this coursework, you need to develop a machine learning (ML) solution to the task of detecting breast cancer from patient diagnostics data.

Dataset

We will use the **Breast Cancer Wisconsin (Diagnostic)** dataset[1]. The dataset is comprised of **569 instances** and **30 features**. You can find more details on the dataset page, including sample code on how to load the dataset up in Python.
The prediction objective is to classify each instance into one of two classes: 'B' (benign) or 'M' (malignant). You may use **any programming language** and **any ML library** of your choice.

Submission Guidelines

- Submit a **PDF-formatted report** accompanied by a working notebook (Jupyter, Colab, or equivalent).
- Your report should clearly describe your methodology, experiments, and findings for each task.
- Marks will be deducted for incomplete or non-functioning code.
- Ensure that your **report ensures easy reproducibility of your findings**, i.e., it should be possible for someone to implement the approach (maybe in a different programming language) without requiring to look at your code.

Tasks (Total Marks: 50)

Develop a Single-Model Logistic Regression Baseline

The dataset doesn't come with a train:test split. **Randomly shuffle the dataset with a fixed seed value**, e.g., "42" (the answer to the ultimate question of life!), and use the **first 70% as train**, **the next 10% as validation**, and the **remaining 20% as evaluation (test) split**.

Q 1 (10 marks). *Train a Logistic Regression classifier with parameters trained on the train split. Clearly specify the settings of your experiments for **reproducibility**:*

- *Hyper-parameter(s) of the logistic classifier, and their value(s).*
- *Training procedure, e.g., the batch size you used etc.*

Develop an Ensemble of Logistic Regression Models

In this task, we are going to investigate if training multiple models and then combining the beliefs from these models (somewhat similar to a Bayesian approach) actually leads to better predictions (it may not; so, it's perfectly okay to report worse results than the baseline).

- Split the training set $\mathcal{T}$ into $n = 5$ partitions $\cup_{i=1}^{n} \mathcal{T}_i = \cup_{i=1}^{n} (\mathcal{X}_i \times \mathcal{Y}_i)$, where $\mathcal{X}_i$ and $\mathcal{Y}_i$ are the corresponding input features and output labels.

[1] `https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic`

- On each split $\mathcal{T}_i$ ($i = 1, \ldots, n$), train a logistic regression model, i.e., $\theta_i : \mathcal{X}_i \mapsto \mathcal{Y}_i$.
- Now, we need to learn a model ϕ that combines the different logistic regression models, $\theta_i, (i = 1, \ldots, n)$ in an optimal way. For this, we will use the **validation split of the data** (i.e., $\mathcal{V}$).
- For each instance in the validation set ($\mathbf{x} \in \mathcal{V}$), we now set its target label as a **one-hot vector** of dimension n (the number of classifier models), where the j-th component is 1 if the predicted value from the j-th model is correct. Two important points:
  - Note that for the validation set, we are allowed to use the true target class values (we assume they are known), i.e., a prediction is correct if $y(\mathbf{x}) = \hat{y}(\mathbf{x}; \theta_j)$.
  - It is quite likely that more than one model will lead to correct predictions, in which case, just **randomly choose one correct model**.
- At the end of this training, you will have $n = 5$ logistic regression classifiers ($\theta_1, \ldots, \theta_n$), and **a single multi-class classifier ensemble** ϕ.

Q 2 (10 marks). *With the mixture model of n binary classifiers trained this way, write down the inference algorithm in your report and implement it.*

***Hint**: You need to use the posterior probabilities obtained from the softmax classifier, ϕ, and also the individual model predictions $\hat{y}(\mathbf{x}; \theta_j)$ from each model θ_j to make the final prediction.*

Compare the Models and Explain the Observations

Q 3 (10 marks). *Present a comparison between the two approaches with the following evaluation measures: i) average **precision** computed over both classes, and ii) **recall** of the **malignant class**.*

Discuss and explain why the ensemble may perform differently from the single model. Consider factors such as variance reduction, model diversity, data partitioning effects, and overfitting/underfitting characteristics.

Q 4 (20 marks). *Report an investigation (with empirical findings) on how you can **calibrate your model towards achieving a higher recall for the malignant class** by changing hyper-parameters of the ensemble approach. Note that you're free to change any hyper-parameters, either of the θ_i's or of ϕ. **Hint**: Clearly describe your experiment settings (what hyper-parameters, how and why they were varied), and report the observations for each hyper-parameter (a plot may be useful) setting explaining the likely reasons for the observations.*