

Machine Learning COIY065H7 2020-2021

Coursework description, guidelines and marking scheme

This coursework is for MSc students only

1. Introduction

This assignment is an integral part of this module and contributes 20% to the overall mark.

Imagine that you are a member of a research group in a company. Your group leader has asked you to understand a neural network training algorithm, implement the method, test it out, explore its behaviour and explain it, and perhaps even suggest ways it might be improved.

To do this assignment you will need a copy of the paper by [Mosca](#), entitled "Training Convolutional Networks with Weight-wise Adaptive Learning Rates", published in the ESANN 2017 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 26-28 April 2017. Note that there is a typo in the paper: as mentioned in the main text of the paper, line 11 of Algorithm-1 should include a division operation, i.e. λ/z_{ij} , not a multiplication as shown in line 11 of Algorithm-1.

You will implement the neural network training algorithm described in the paper, called WAME, and you will test it on a real-world data set. You have to describe how the method works, adjust its free parameters, discuss your experimental set up and explain your experiments. Lastly, you have to evaluate how the algorithm solves the problem, and discuss your results and their significance.

By doing this coursework you will get experience with implementing, running, adapting, and evaluating machine learning methods on real data. You will need to write, reuse or change code, run it on some data, make some figures, read a few background papers, present your results, and write a report describing the problem you tackled, the machine learning algorithm you used and the results you obtained.

For your experimental study you must use the following real-world data set

[Adult dataset](#): the aim is to predict whether income exceeds \$50K/yr based on census data. Also known as "Census Income" dataset. For more details, consult:
<https://archive.ics.uci.edu/ml/datasets/adult> or an article that used the dataset:

R. Kohavi, Scaling Up the Accuracy of Naive-Bayes Classifiers: a Decision-Tree Hybrid, Proc. of the 2nd International Conference on Knowledge Discovery and Data Mining, 1996. The paper can be downloaded from <http://robotics.stanford.edu/~ronnyk/nbtrees.pdf>.

You should use the above data set. You cannot choose any data you wish.

You can use any programming language or software library for this assignment. In the labs, we use MATLAB which provides well tested functions for neural networks design that are appropriate for the UCI dataset mentioned above. It is not required to program everything from scratch but if you reuse code or libraries you have to cite the original source, otherwise you might be accused for [plagiarism](#).

The assignment is further explained in [Sections 2](#) below. [Section 3](#) of this document gives you an example of how to structure your report and explains the marking scheme. [Section 4](#) presents the deadlines and submission instructions. [Section 5](#) explains the penalties for late submissions, and [Section 6](#) explains how the College deals with plagiarism. [Section 7](#) and [Section 8](#) provide additional information on learning resources and referencing.

2. Implementation and experimentation

You can do your WAME implementation in MATLAB, write your own code or build on a package/library from the internet. You are not tested on programming so the coding style does not have to be perfect and your code does not have to be optimal but it should obviously work correctly. Since the focus is on implementing WAME, I wouldn't recommend implementing all the methods required for training neural networks, e.g. backpropagation, derivatives calculation etc., from scratch unless you are very experienced with Java, C++, Python or some other programming language or platform. No matter what you do/use, make sure that all sources and code taken by others or the internet are cited properly in your Report; otherwise, you may be accused of [plagiarism](#).

Some packages provide techniques for determining the optimal structures of machine learning models (e.g. the model architecture or the model hyperparameters) automatically as part of the training. In that case, instead of performing experimental tests varying the number of free parameters of the model, these techniques can be used to find the appropriate structure for your model. Still some of these methods may have their own parameters, which require fine tuning.

Note that the performance results of your approach would be more meaningful, if a validation technique is used, such as k-fold cross validation ($k=7$ or $k=10$ is typically used), or leave-one-out cross validation, or some form of Monte Carlo simulation. Lastly, the use of regularisation, provided in some software packages and in Matlab, normally helps to get better results.

The results of your experiments should be stored in ASCII format, in a Jupyter notebook, or in notebook documents produced by other web-based interactive computational environments, specifying whether the result is from the training or the testing phase, and should be submitted together with your report (Moodle will allow to submit additional files; up to 5 files in total can be submitted). Check that these files can be opened and read correctly. Results should be presented using figures and tables and discussed in your Report, i.e. it is not enough just to submit a python notebook or files with results- these are not accepted as a Report submission.

3. Assignment outline and marking scheme

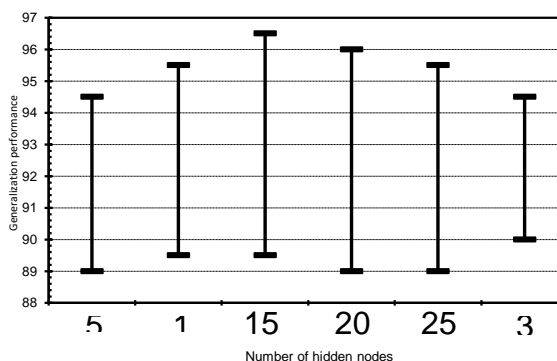
Your work will be presented in a Report (notebook documents are not accepted as a Report). It is important that your Report is properly structured. Sections like the ones shown below could be included in your report to ensure good coverage of the topic. A number of 2000-3000 words are expected to cover in sufficient depth all aspects of the assignment, but **our marking is not based on the number of words used in the Report**. What primarily matters is that you describe your design of the training algorithm, justify any choices you made, explain how things work, and make the model work with data. Also, you will need to provide insight on how to (pre)process the data before feeding into the model and do the training (if necessary), how to debug the learned model (not only the training algorithm), how to measure model performance and demonstrate its significance. **You are not just being marked on how good the results of your training algorithm or neural model are. However, when awarding marks, the work described in the report is marked in comparison to the work of other students. The following marking scheme is used.**

1. Methodology, design and technical contribution (50% of the mark): the appropriate use and sophistication of design methods, overall methodology and WAME implementation is marked here

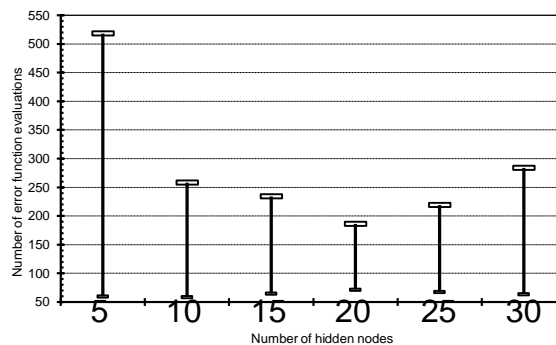
- 1.1 This part should normally describe clearly the approach used in your design and implementation and any relevant parameters (e.g. for neural networks models this includes number of hidden nodes, layers, type of activation functions, training algorithm parameters etc). If you are using a particular library or tool, you still need to describe how the methods/functions that you are using operate and what parameter values are used. Citing the library, tool, etc., and just listing the library functions is not enough to get a good mark.
- 1.2 This part should describe and justify any special techniques/methods or parameters used in the stages of your methodology, e.g. in data preprocessing, in WAME initialisation or during training. Also, this part should describe and justify any techniques for normalisation or missing data, or other pre-processing or balancing methods, and whether you have used some form of cross-validation, regularisation, or weight decay, providing details of the particular method. Citing the library, tool, etc., or simply mentioning the library functions is not enough to get a high mark.
- 1.3 This part should describe your technical contribution: if your code fully runs or not and how you have tested the code (especially important if your code does not run- it makes sense to show that you have made a serious attempt to identify any issues); if the code is based on existing code and whether you have made any effort to combine different components; whether a novel implementation of the WAME has been produced and/or you have produced your own modifications. If the implementation has problems/bugs, it does not automatically mean that you have failed the coursework.

2. Experiments, findings and discussion (40% of the mark): the experimental design and the investigation's systematicity are marked here. You must present and discuss your experimental design and results. You are expected to run several experiments, explore the behaviour of the WAME training algorithm under different conditions and calculate basic statistics to summarise performance in training and testing. You can explore the behaviour of other neural network training algorithms if you want or other machine learning methods. That work will enhance the experimental study, but the core of the study is expected to be on WAME performance and parameters, not on studying methods already available in the packages/libraries. You also need to discuss the significance of your findings or comparisons that you have made. Your report must include at least two figures which graphically illustrate quantitative aspects of your results, such as training/testing errors, performance metrics for sets of learned parameters, algorithm outputs, descriptive statistics, etc.

For example, in this part you can use Excel or other packages to provide charts - like the figure below, which uses error bars (Box and Whisker Charts in Excel), to show the performance of your trained model in terms of generalisation. For example, the figure below shows generalisation with respect to number of hidden nodes used in a neural network-based solution. Alternatively, one could use tables to provide the same information by giving for each number of hidden nodes the average value, the minimum value, and the maximum value of generalisation performance (in percentage of successfully recognised patterns) in the tests.



You could also discuss the cost of the computations, e.g. referring to the number of training iterations required or the number of error function evaluations (see figure below for the neural-network based solution discussed above). In the same way, you can also present and discuss other aspects of the training phase, e.g. impact of hyperparameters, folds, balancing methods etc.



In machine learning, experimental comparison/evaluation is presented for a few methods and their parameters. Overall results are also presented in tables like the one below that shows average performance of two models, trained using two different training algorithms, on the same test set. In this case, performance is shown in terms of recognition success per class as well as average classification success. Confusion matrices can also be used.

Method	Class 1 (%)	Class 2 (%)	Average success (%)
Algorithm 1- trained Model	83	96	93
Algorithm 2- trained Model	73	93	88

3. Conclusions (10% of the mark)

- 3.1 Provide an honest and justified overview/summary of your work and a critical view of the findings.
- 3.2 Identify areas for improvement; discuss what you could have done better (particularly important if you failed some of your targets or your results are not as expected).

4. Bibliography: this is not marked directly but is necessary as it supports the justification of your methodology and methods and the claims/arguments in your report. Make sure that all sources are cited in the text of the report and are also listed in the bibliography section- this way you don't get in any trouble with plagiarism detection software.

Provide a list of the bibliographical/web sources you used. Include publication details and all information necessary to access the online resources. Sources should be cited in the text by (Author name, year) and appear in the references list in alphabetical order by Author's last name. This also applies to websites, e.g. an online article/webpage should be listed in your references; for example:

(MLOSS, 2011). *Machine learning open source repository*. Available online at <http://mloss.org/>

NOTE: use of any text or code (even open source code) taken from other sources should be clearly identified and referenced in your report to avoid plagiarism (see [Section 6](#)). If you are unsure on which parts of your code needs appropriate referencing do consult the module lecturer.

4. Deadlines and submission instructions

Submission is only through Moodle and consists of the submission of a Report, and data files with your results and code as appendices (if existing Matlab toolboxes have

been used or libraries, these should be mentioned in the Report). Important parts of your code can be included and discussed in the report but a **Notebook document, copy of the Matlab editor etc. is not accepted as a Report submission. The preferred format for the Report is Word document but also PDF or RTF are accepted with data and code embedded in the file. You report and the code will be tested for plagiarism.**

We are unfortunately unable to accept and mark reports that contain text embedded as an image into the document. These documents cannot go through the plagiarism detection system and will not be marked.

Make sure you are familiar with Moodle and able to upload your files (for example you could test the system by uploading a test file). **Hardcopy versions of the report/data/code files will not be accepted.**

You should upload on Moodle the completed assignment by **April 27th, 2021 at 11:00pm** (this is Moodle time not your PC's time. In case you are planning to upload your files whilst at a remote location make sure you check Moodle's time and take into account time zone differences).

Your files should be named according to your last name.

Your Report must have a cover page!

The first page of your report **MUST** have the following information:

Module title and code: Machine Learning- COIY065H7

Name: your first name and last name

Emails: please provide your College email and the email you use- if different from your College email

Your Report should have an Appendix with a description of data files and code submitted; when these are not included in the Report, as appendices, but are submitted separately. Any specific instructions on software use should also be included in an Appendix of the report, entitled "Instructions for using the code".

It is your responsibility to ensure that files transferred from your own machines are in the correct format and that any programs execute as intended on Department's systems prior to the submission date.

Each piece of submitted work **MUST** also have a page entitled "Academic Declaration" by the author that certifies that the author has read and understood the sections of plagiarism in the document <http://www.bbk.ac.uk/mybirkbeck/services/rules/Assessment%20Offences.pdf> that describes College's Policy on assessment offences. Confirm that the work is your own, with the work of others fully acknowledged. Submissions must also be accompanied by a declaration giving us permission to submit your report to the plagiarism testing database that the College is using.

Reports without a Declaration form are not considered as completed assignments and are not marked.

The Academic Declaration should read as follows: "I have read and understood the sections of plagiarism in the College Policy on assessment offences and confirm that the work is my own, with the work of others clearly acknowledged. I give my permission to submit my report to the plagiarism testing database that the College is using and test it using plagiarism detection software, search engines or meta-searching software."

You should note that all original material is retained by the Department for reference by internal and external examiners when moderating and standardising the overall marks after the end of the module. We are aiming to provide grades and feedback through Moodle by June

2nd, 2021. Original marks awarded will be shown on Moodle, i.e. any penalty or capping will be applied at a later stage after examining any mitigating circumstances claims.

Those who would like to get some early feedback on their assignment before submitting the completed work, they can email <gmagoulas@dcs.bbk.ac.uk> their questions or send a draft of their report for comments. This should be done **before the College's Easter closing date** as we may not be able to respond to messages in a timely manner after this date.

5. Late coursework

It is our policy to accept and mark late submissions of coursework. You do not need to negotiate new deadlines and there is no need to obtain prior consent of the module lecturer.

The **last day the system will accept a late submission for this module is May 11th, 2021 at 11:00pm** (this is Moodle time not your PC's time. In case you are planning to upload your files whilst at a remote location make sure you check the Blackboard time and take into account time zone differences). **May 11th, 2021 at 11:00pm is the absolute cut-off deadline for coursework submission. Work from this point on will be treated as a non-submission.**

According to the College's regulation, penalty applies on late submissions as follows:

- **Work that is submitted up to 7 calendar days late will be subject to a late penalty representing a 10% deduction of the original mark awarded.**
- **Any work that is submitted more than 7 calendar days late will receive a late capped penalty mark of 50%.**
- **Work submitted after the absolute cut-off deadline (May 11th 2021) gets zero (0) marks.**

If you believe you have good cause to be excused the penalty for late submission of your coursework, you must make a written request using a mitigating circumstances application form and attach any evidence. Your form should be emailed to the Postgraduate Programme Administrator (with a carbon copy to the module lecturer and the Programme Director) as soon as possible, ideally that is by the cut-off deadline. This email does not need to be submitted at the same time as you submit the coursework itself but **MUST be submitted by May 17th, 2021.**

Even if the personal circumstances that prevented you from submitting the coursework by the last day (i.e. May 11th, 2021) are extreme, **the Department will not accept coursework after this date.** We will, naturally, be very sympathetic, and the MSc Programme Director will be happy to discuss ways in which you can proceed with your studies, but please do not ask us to accept coursework after this date; we will not be able to as there is a College-wide procedure for managing late submissions and extenuating circumstances in student assessment. As soon as you know that you will not be able to meet the deadline, it will be useful for you to inform the module lecturer. They will be able to advise you on how best to proceed. Another person to speak to, particularly if the problem is serious, is the MSc Programme Director. You will then have the opportunity to discuss various options as to how best to continue your studies.

Further details concerning the rules and regulations with regard to all matters concerning assessment (which naturally includes coursework), you should consult College Regulations at <http://www.bbk.ac.uk/mybirkbeck/services/rules>. **Please check your MSc programme 2020/21 booklet for any special rules governing Late Submissions and consideration of Mitigating Circumstances, and the official Policy for Mitigating Circumstances at the College's website <http://www.bbk.ac.uk/registry/policies/policies-2020-21> .**

6. Plagiarism

The College defines plagiarism as “copying a whole or substantial parts of a paper from a source text (e.g. a web site, journal article, book or encyclopedia), without proper acknowledgement; paraphrasing of another’s piece of work closely, with minor changes but with the essential meaning, form and/or progression of ideas maintained; piecing together sections of the work of others into a new whole; procuring a paper from a company or essay bank (including Internet sites); submitting another student’s work, with or without that student’s knowledge; submitting a paper written by someone else (e.g. a peer or relative), and passing it off as one’s own; representing a piece of joint or group work as one’s own”.

The College considers plagiarism a serious offence, and as such it warrants disciplinary action. This is particularly important in assessed pieces of work where the plagiarism goes so far as to dishonestly claim credit for ideas that have been taken from someone else.

Each piece of submitted work MUST have an “Academic Declaration” form signed by the student which certifies that the students have read and understood the sections of plagiarism in the College Regulation and confirm that the work is their own, with the work of others fully acknowledged. Submissions must be also accompanied by a declaration giving us permission to submit coursework to a plagiarism testing database that the College is subscribed.

If you submit work without acknowledgement or reference of other students (or other people), then this is one of the most serious forms of plagiarism. When you wish to include material that is not the result of your own efforts alone, **you should make a reference to their contribution, just as if that were a published piece of work.** You should put a clear acknowledgement (either in the text itself, or as a footnote) identifying the students that you have worked with, and the contribution that they have made to your submission.

7. Referencing

References include the full bibliographic information about the source, such as the author(s)'s name(s), date of publication, title of work, place of publication, and publisher. This information is usually given in the section called Reference List or Bibliography at the end of the text. The key principle is that you should give enough information to allow another person to find the source for themselves.

Here are some examples using the Harvard referencing system:

[when you are referring to a book]

Lewin, K., 1951. *Field Theory in Social Science*. New York: Harper and Row.

[when you are referring to a chapter in a book, where 'ed.' means editor, and 'edn.' means 'edition']

Piaget, J., 1970. Piaget's theory. In: P. Smith, ed., *Handbook of child psychology*. 3rd edn. New York: Wiley, 1970, pp. 34-76.

[when you are referring to a journal article]

Holmqvist, M., 2003. A Dynamic Model of Intra- and Interorganizational Learning. *Organization Studies*, 24(1), 95-123.

[when you are referring to a webpage]

W3C, Web Accessibility Guidelines and Techniques, available online at <http://www.w3.org/WAI/guid-tech.html>. Last accessed 12/02/2015.

Independent of their type (e.g. book, article, webpage), all references are included at the end of a document in alphabetical order starting from the author's name as in the example above.

8. Useful resources

Here are some resources on plagiarism, study skills, and referencing that can help you to better manage your project and avoid plagiarism.

Avoiding Plagiarism

<https://moodle.bbk.ac.uk/enrol/index.php?id=45>

Referencing Systems

Harvard guide to citing references Available to online at:

http://www.open.ac.uk/libraryservices/documents/Harvard_citation_hlp.pdf

Study Skills

<https://moodle.bbk.ac.uk/enrol/index.php?id=12219>

If you need guidance relating to your studies and identifying needs for additional support, please contact your personal tutor. The personal tutor may be also able to help if you have other problems that may be relevant to your Programme of study.

The School or College also provide advice sessions on study skills, such as planning the workload, writing, reading and exam preparation; free workshops on critical writing, essay writing, examination essays, examination revision, learning strategies and memory techniques; help with the English language, help with coursework and other written assessment, subject specific preparation and study skills:

<http://www.bbk.ac.uk/student-services/learning-development> .

Lastly, the School's Learning Development Tutors are there to support you in your studies and offer help and advice on a wide range of study skills:

<http://www.bbk.ac.uk/business/current-students/learning-co-ordinators> .