ECE5268 Theory of Neural Networks (Spring 2021)
Mini-Project #1 (ver. 1.0)

Dr. Georgios C. Anagnostopoulos

Tuesday 2nd March, 2021

1 Objectives

The objective of Mini-Project (MP) I is to assess students' understanding in a variety of aspects pertaining to Ordinary Least Squares (OLS).

Before putting together your work, carefully review the preparation guidelines (Section 3) and submission instructions (Section 4) that are provided.

2 Tasks

The credit assignments for each task are as follows:

Problem	Part	Points
1	(a)	5
	(b)	10
	(c)	5
	(d)	10
	(e)	10
2	(a)	10
	(b)	10
	(c)	10
	(d)	10
3	(a)	10
	(b)	10
Total		100

Problem 1. [Prob-LinReg03] Assume that we would like to fit a linear regression model using a training set $\mathcal{T} \triangleq \left\{\left(\mathbf{x}_n \in \mathbb{R}^D, y_n \in \mathbb{R}\right)\right\}_{n=1}^N$ and squared loss (so, it is an OLS problem). Let $\mathbf{X} \in \mathbb{R}^{N \times D}$ be the problem's design matrix, whose rows consists of the input samples $\mathbf{x}_n$ and $\mathbf{y} \in \mathbb{R}^N$ be the vector of the corresponding, paired target outputs y_n. Furthermore, let $\widetilde{\mathbf{X}} \triangleq [\mathbf{X} \quad \mathbf{1}_N] \in \mathbb{R}^{N \times (D+1)}$, whose last column is a column of ones, be the problem's augmented design matrix. Finally, let the augmented model parameter vector be $\widetilde{\mathbf{w}} \triangleq [\mathbf{w}^T \quad b]^T \in \mathbb{R}^{D+1}$, so that the model's predicted output $\hat{y}$ for input $\mathbf{x}$ is given as $\hat{y} = \mathbf{x}^T\mathbf{w} + b$. Training the linear regression model amount to finding

$$\widetilde{w}^* \triangleq \arg\min_{\widetilde{\mathbf{w}} \in \mathbb{R}^{D+1}} \underbrace{\frac{1}{N}\left\|\widetilde{\mathbf{X}}\widetilde{\mathbf{w}} - \mathbf{y}\right\|_2^2}_{\mathrm{MSE}_{\mathcal{T}}(\widetilde{\mathbf{w}}) \triangleq} \tag{1}$$

where $\mathrm{MSE}_{\mathcal{T}}(\widetilde{\mathbf{w}}) = \mathrm{MSE}_{\mathcal{T}}(\mathbf{w}, b)$ is the Mean Squared Error (MSE) as evaluated on the training set $\mathcal{T}$ and $\widetilde{\mathbf{w}}^*$, which consists of $\mathbf{w}^*$ and b^*, is the parameter vector that minimizes it. Finally, define the following quantities:

$$\overline{\mathbf{x}} \triangleq \frac{1}{N}\mathbf{X}^T\mathbf{1}_N = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}_n \tag{2}$$

$$\overline{y} \triangleq \frac{1}{N}\mathbf{y}^T\mathbf{1}_N = \frac{1}{N}\sum_{n=1}^{N}y_n \tag{3}$$

$$\check{\mathbf{X}} \triangleq \mathbf{P}_N\mathbf{X} = \mathbf{X} - \mathbf{1}_N\overline{\mathbf{x}}^T \tag{4}$$

$$\check{\mathbf{y}} \triangleq \mathbf{P}_N\mathbf{y} = \mathbf{y} - \overline{y}\mathbf{1}_N \tag{5}$$

where

$$\mathbf{P}_N \triangleq \mathbf{I}_N - \frac{1}{N}\mathbf{1}_N\mathbf{1}_N^T \tag{6}$$

is the $N \times N$ *centering matrix*.

Part (a):
Show that the optimal (minimizing the training MSE) hyper-plane of a linear regression model with an intercept always passes through the point $(\bar{\mathbf{x}}, \bar{y})$, *i.e.*, it holds that

$$\bar{y} = \bar{\mathbf{x}}^T\mathbf{w}^* + b^* \tag{7}$$

In order to do so, use the fact that $(\mathbf{w}^*, b^*)$ is a minimizer of the traning MSE and, therefore,

$$\left.\frac{\partial \mathrm{MSE}_{\mathcal{T}}(\mathbf{w}, b)}{\partial b}\right|_{\mathbf{w}=\mathbf{w}^*, b=b^*} = 0 \tag{8}$$

Part (b):
Use the result of (7), *i.e.*, the fact that the optimal value of the intercept term is $b = \overline{y} - \overline{\mathbf{x}}^T\mathbf{w}$, to show that

$$\mathbf{w}^* = \arg\min_{\mathbf{w} \in \mathbb{R}^D} \mathrm{MSE}_{\mathcal{T}}(\mathbf{w}, \overline{y} - \overline{\mathbf{x}}^T\mathbf{w}) = \arg\min_{\mathbf{w} \in \mathbb{R}^D} \frac{1}{N}\left\|\check{\mathbf{X}}\mathbf{w} - \check{\mathbf{y}}\right\|_2^2 \tag{9}$$

Part (c):
Use (9) to show that

$$\mathbf{C}_{x,x}\mathbf{w}^* = \mathbf{c}_{x,y} \tag{10}$$

where

$$\mathbf{C}_{x,x} \triangleq \frac{1}{N}\check{\mathbf{X}}^T\check{\mathbf{X}} = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{x}_n - \overline{\mathbf{x}})(\mathbf{x}_n - \overline{\mathbf{x}})^T \tag{11}$$

$$\mathbf{c}_{x,y} \triangleq \frac{1}{N}\check{\mathbf{X}}^T\check{\mathbf{y}} = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{x}_n - \overline{\mathbf{x}})(y_n - \overline{y}) \tag{12}$$

Above, $\mathbf{C}_{x,x}$ is the sample covariance matrix of the input samples and $\mathbf{c}_{x,y}$ is the sample covariance vector of the inputs vs. the target outputs.

Part (d):
Assuming that $\mathbf{C}_x$ is invertible, show that

$$\mathrm{MSE}_{\mathcal{T}}(\mathbf{w}^*, b^*) = c_{y,y} - \|\mathbf{c}_{x,y}\|^2_{\mathbf{C}_x^{-1}} \tag{13}$$

where

$$c_{y,y} \triangleq \frac{1}{N}\|\check{\mathbf{y}}\|_2^2 = \frac{1}{N}\sum_{n=1}^{N}(y_n - \overline{y})^2 \tag{14}$$

is the sample variance of the target outputs.

Part (e):
The *coefficient of determination* R^2 ("R squared") of a linear regression model that uses an intercept term b and that is fitted to a training set $\mathcal{T}$ is defined as

$$R^2 \triangleq \begin{cases} 1 & \text{if} \quad y_n = \text{const} \quad \text{for all } n \\ 1 - \frac{\min\limits_{\mathbf{w}\in\mathbb{R}^{\mathbb{D}}, b\in\mathbb{R}} \mathrm{MSE}_{\mathcal{T}}(\mathbf{w}, b)}{\min\limits_{b\in\mathbb{R}} \mathrm{MSE}_{\mathcal{T}}(\mathbf{0}, b)} & \text{otherwise} \end{cases} \tag{15}$$

First, show that

$$\min_{b\in\mathbb{R}} \mathrm{MSE}_{\mathcal{T}}(\mathbf{0}, b) = c_{y,y} \tag{16}$$

Next, show that, for the case, where not all y_n's equal the same constant value,

$$R^2 = \frac{\|\mathbf{c}_{x,y}\|^2_{\mathbf{C}_{x,x}^{-1}}}{c_{y,y}} \tag{17}$$

and that

$$0 \leq R^2 \leq 1 \tag{18}$$

Comments:

1. From (9), we see that the optimal intercept term b^* is such that it zero-means the inputs and target outputs and $\mathbf{w}^*$ can be obtained by solving an OLS problem without an intercept term.

2. From (12) and (13), we see that, if the centered inputs and centered targets are uncorrelated ($\mathbf{c}_{x,y} = \mathbf{0}$), then $\mathbf{w}^* = \mathbf{0}$, the optimal linear predictor becomes $\hat{y}^* = b^* = \overline{y}$ and the optimal training MSE achieves its maximum value of $\frac{1}{N}\|\check{\mathbf{y}}\|_2^2$.

Problem 2. [Prob-LinReg07] Consider the dataset provided in `autompg_dataset.csv`, which was derived from the Auto MPG Data Set dataset hosted at the UCI ML Repository by removing samples with missing feature values. We are going to use this dataset to predict a car's miles-per-gallon fuel consumption `mpg` using OLS regression based on a number of car characteristics.

First, split the available data into a training set (first 200 samples), a validation set (the next 100 samples) and a test set (remaining 92 samples).

Part (a):
By minimizing the training MSE, fit a linear regression model with intercept (let's call it *Model1*) that predicts `mpg` based on the `horsepower` and `weight` input features. Report (i) the fitted model's prediction equation and (ii) the optimal training MSE value achieved. Does the prediction equation make intuitive sense or not? Justify your answer.

Part (b):
Produce a plot showing the fitted model's response surface (it will be a plane) superimposed on the scatter plot of the training data (`horsepower`, `weight`, `mpg`). Comment on the obtained plot.

Part (c):
Repeat part (a), but, this time, use all available input features. Let's call this model *Model2*. Compare the obtained results to the ones of part (a).

Part (d):
Select the best of *Model1* and *Model2* based on the validation MSE. Finally, for the best model, report its test MSE. Provide pertinent commentary on your findings.

Problem 3. [Prob-SGD05] Consider the OLS regression task of Problem 2 of linearly predicting `mpg` based on the `horsepower` and `weight` input features. In this problem we will solve this task via the mini-batch Stochastic Gradient Descent (SGD) minimization method with a $\mathcal{O}(1/k)$-decreasing learning rate.

Have your SGD implementation always start at $\widetilde{\mathbf{w}}^{(0)} = \mathbf{0}$, run for a maximum of 1000 epochs and report the following for every epoch: (i) the $\log_{10}$ value of the training MSE, (ii) $\log_{10}$ value of the validation MSE. The implementation should track and return as the problem's solution the model parameters that achieve the lowest validation MSE encountered.

Part (a):
For a batch size of 1, produce a plot of the $\log_{10}$ value of the training MSE and the $\log_{10}$ value of the validation MSE versus epochs. Experiment with different learning rate parameters $\eta_{\max} > 0$ and $\beta > 0$, so that $\log_{10}$ training MSE is mostly decreasing with increasing epochs. Report the fitted model obtained in this manner and its validation MSE. Compare them to the corresponding findings of Problem 2 and supply pertinent comments.

Part (b):
Repeat the previous part, but, now, for a batch size of 20. How do these runs and their results compare to the ones of the previous part?

3 Preparation Guidelines

Below are some general guidelines that you are asked to follow, when compiling a project report. These guidelines greatly facilitate your work's assessment by a grader and, at the same time, aim at helping you sidestep some major pitfalls that would prevent you from receiving the maximum credit for your work.

- **Assignment Statements:** Before attempting to address a particular assignment, ensure that you completely understand what is asked from you to perform and/or to produce. When in doubt, ask your instructional staff for clarifications! Also, make sure you did not omit your response to any of the parts that you have attempted. Finally, make sure that it is crystal clear, which response corresponds to which assignement/part.

- **Derivations & Proofs:** If you provide handwritten derivations and/or proofs, make sure you use your best handwriting. Each derivation should have a logical and organized flow, so that it is easy to follow and verify.

- **Code & Data:** The code that you author should be as well-organized as possible and amply commented. This is very useful for assessing your work, as well as for you, while you are debugging/or modifying it, or when you have to go back to it in the near future. **Caution:** You are not allowed to use any code that you have not produced without having/obtaining explicit prior permission, in which case the source(s) you have obtained this code from must be clearly indicated via comments inside your code as well in your report. You are deemed to be plagiarizing, if you fail to do so, which may have dire consequences. Finally, if a task asks you to generate data, keep them organized in a separate folder and document, *e.g.*, in a text file, the specifics of how they were generated.

- **Figures, Plots & Tables:** Plots should have their axes labeled and, if featuring several visual elements such as curves or types of points on the same graph, an appropriate legend should be used. Whether figures or tables, each one of these elements should feature a caption with sufficient information on what is being displayed and how were these results obtained (*e.g.*, under what experimental conditions or settings, etc.). You should ask yourself the question: if someone comes across it, will they understand what is being depicted? Apart from a concise description, major, relevant conclusions stemming from the display should also be included in the caption text. Keep in mind, that when describing a figure, do not describe only what is shown, but describe what it means.

- **Observations, Comments & Conclusions:** When stating observations about a particular result, do not stop at the obvious that anyone can notice (*e.g.*, "*... we see that the curve is increasing.*"). Instead, assess whether the result is expected, either by theory or intuition (*e.g.*, "*... This is as expected, because X is the integral of ...*"), or, if it is unexpected, offer a convincing reasoning behind it (*e.g.*, "*... We expected a decreasing curve ... All points to that I must have not been calculating X correctly ...*"). The latter is more preferable (*i.e.*, expect partial credit) than stopping at the obvious, which happens to be wrong (*i.e.*, do not expect partial credit). Next, descriptions and comments on results should be sufficient. Be concise, but complete. Finally, conclusions that you draw must be well-justified; vacuous conclusions will be swiftly discounted.

4 Submission Instructions

Kindly adhere to the conventions and submission instructions outlined below. Deviations from what is described here may cause unnecessary delays, costly oversights and immense frustrations related to the assessment of your hard work.

Important: Your MP work needs to be submitted via Canvas. Please do not email it to the instructional staff.

First, store all your project deliverables in a folder named `lastname_mpX`, where `lastname` should be your last name and `X` should be the number of the MP, like 1, 2, etc. The folder name should be all lower case. For example, Anagnostopoulos' folder for MP 1 would be named: `anagnostopoulos_mp1`.

Secondly, your `lastname_mpX` folder should have the following contents:

- A signed & dated copy of the Work Origination Certification page in Adobe PDF format. You can either scan such a page after you complete, date and sign it, or do so electronically, as long as your signature is not typed. If this page is missing from your report, your MP work will not be considered for assessment (grading) and will be assigned a default total score of 0/100.
- An Adobe PDF document named **lastname_report.pdf**, where, again, "lastname" should be replaced by your last name in all lower case, *e.g.*, *anagnostopoulos_report.pdf*. This document should contain your entire Mini-Project report as a single document. This will be the document that will be graded. Also, here are some important things to adhere by:
  - Your responses to the project's assignments should be given in the order in which these assignments appear in the MP's statement. If you did not attempt a particular part, list it in its expected order and state that you have not attempted it as your response.
  - For assignments that require you to show analytical work (*e.g.*, a derivation/proof), you are not obliged to typeset it. While it would be nice to do so, such effort may turn out to be quite time-consuming. Instead, you can scan your work into an image, as long as it is legible and well organized with a clear logical flow. When scanning your hand-written work, use a relatively low-resolution (DPI) setting, so your resulting PDF document does not become too big in size, which may prevent you from uploading your work to Canvas. Use a scanner, not a photo taken by a mobile device.
- A folder named `src`, which should contain all your code (*e.g.*, MATLAB or Python scripts, Jupyter Notebooks, etc.) that you authored and used for producing your results and the data sets that you created for this MP, if applicable. It is best, if you named your scripts according to the task and or task/part pair they produce results for.
- An optional folder named `docs`, in which you can include a MS Word version of your report and other ancillary material connected in one way or another to your MP report.

WORK ORIGINATION CERTIFICATION

By submitting this document, I, ____, the author of this deliverable, certify that

1. I have reviewed and understood FIT Student Handbook's policies on Academic Honesty, which discusses issues such as plagiarism, cheating and miscellaneous misconduct.
2. The content of this MP submission reflects my personal work and, in cases it is not, the source(s) of the relevant material has/have been appropriately acknowledged after it has been first approved by the course's instructional staff.
3. In preparing and compiling all this material, I have not collaborated with anyone and I have not received any type of help from anyone but the course's instructional staff.

Signature ____ Date _______________