ECE 5268 - Theory of Neural Networks (Spring 2017)
Mini-Project #1

Dr. Georgios C. Anagnostopoulos*

March 2017

1 Objectives

The objective of Mini-Project I is to expose the students to (i) sequential Bayesian Linear Regression (BLR), (ii) the application of Ridge Regression (RR), (iii) the application of Radial Basis Function (RBF) Artificial Neural Networks (ANNs) and (iv) practical aspects of model selection via validation.

As usual, standard preparation guidelines (Section 3) and submission instructions (Section 4) are provided, which the students are expected to strictly adhere to. Finally, at the end of this document, a few, possibly helpful, references can be found.

2 Assignments

Task 1. [35 total points]

This task refers to the recursive (sequential) calculation of the Maximum A Posteriori (MAP) estimate $\hat{\mathbf{w}}_{MAP}$ of $\mathbf{w}$ in the context of single-output BLR. Consider the scenario, where input/output pairs become available in a piece-meal fashion, *i.e.*, the training set grows as time goes by. Obviously, one could augment the design matrix $\mathbf{X}$ by extra rows and the target vector $\mathbf{t}$ by an equal number of extra elements and recompute $\hat{\mathbf{w}}_{MAP}$. However, (i) this seems to be a waste of computations, as we "throw away" the previous estimate instead of taking advantage of it and (ii) in some real-time applications (*e.g.*, tracking, forecasting, etc.) the cost of recomputing the estimate "from scratch" is prohibitive.

For Part (a) through Part (d), assume a BLR problem with data generation process $t_n = \mathbf{x}_n^T\mathbf{w} + e_n$ for $n = 1, \ldots, N$, where $e_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \sigma^2\right)$ with known value of $\sigma^2 > 0$ and prior distribution $\mathbf{w} \sim \mathcal{N}_D\left(\mathbf{w}_o, s^2\mathbf{I}\right)$ with some $s^2 > 0$. Denote as $\mathbf{C}_{\mathbf{w}|\mathbf{t}}(N)$ and $\hat{\mathbf{w}}_{MAP}(N)$ the posterior covariance matrix of $\mathbf{w}$ and the MAP estimate of $\mathbf{w}$ for the N available data. Assume now that a new sample pair $(\mathbf{x}_{N+1}, t_{N+1})$ becomes available.

For Part (e) and Part (f), consider the data generation process $t_n = w_1 + w_2 x_n + e_n$, where $e_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \frac{1}{4}\right)$ and assume a-priori that $\mathbf{w} \sim \mathcal{N}_2\left(\mathbf{0}, s^2\mathbf{I}\right)$ for a suitable value of $s^2 > 0$, which you may choose to your liking. Use the `task1` dataset, whose first column is x and second column is t.

(a) **[10 points]** If $\mathbf{C}_{\mathbf{w}|\mathbf{t}}(N+1)$ stands for the (updated) posterior covariance matrix of $\mathbf{w}$ computed based on the $N+1$ samples, show that

$$\mathbf{C}_{\mathbf{w}|\mathbf{t}}^{-1}(N+1) = \mathbf{C}_{\mathbf{w}|\mathbf{t}}^{-1}(N) + \frac{1}{\sigma^2}\mathbf{x}_{N+1}\mathbf{x}_{N+1}^T \tag{1}$$

*georgio@fit.edu

(b) **[5 points]** Using Equation (1) and the special Sherman-Morisson-Woodbury identity, show that

$$\mathbf{C}_{\mathbf{w}|\mathbf{t}}(N+1) = \mathbf{G}(N+1)\mathbf{C}_{\mathbf{w}|\mathbf{t}}(N) \tag{2}$$

where we define

$$\mathbf{G}(N+1) \triangleq \mathbf{I} - \frac{\mathbf{C}_{\mathbf{w}|\mathbf{t}}(N)\mathbf{x}_{N+1}\mathbf{x}_{N+1}^T}{\sigma^2 + \mathbf{x}_{N+1}^T\mathbf{C}_{\mathbf{w}|\mathbf{t}}(N)\mathbf{x}_{N+1}} \tag{3}$$

(c) **[5 points]** Now, based on Equation (2) and Equation (3) show that

$$\hat{\mathbf{w}}_{MAP}(N+1) = \mathbf{G}(N+1)\hat{\mathbf{w}}_{MAP}(N) + \frac{t_{N+1}\mathbf{C}_{\mathbf{w}|\mathbf{t}}(N)\mathbf{x}_{N+1}}{\sigma^2 + \mathbf{x}_{N+1}^T\mathbf{C}_{\mathbf{w}|\mathbf{t}}(N)\mathbf{x}_{N+1}} \tag{4}$$

where $\hat{\mathbf{w}}_{MAP}(N+1)$ is the (updated) MAP estimate of $\mathbf{w}$ based on the $N+1$ data.

(d) **[5 points]** Equation (1) through Equation (4) form the update equations for sequential BLR, which allows for computing of posteriors quantities by augmenting the data set one sample at a time. How should $\mathbf{C}_{\mathbf{w}|\mathbf{t}}(0)$ and $\hat{\mathbf{w}}_{MAP}(0)$ be initialized, before any datum is available ($N = 0$), so that the sequential algorithm yields the correct result for any $N \geq 1$? Justify your answers by showing that, indeed, you get a correct algorithm.

(e) **[5 points]** By using the sequential BLR algorithm for $N = 0, 1, \ldots, 1000$ on the `task1` data set, produce a MATLAB® animation (see [2] for how you can go about it) in the way described below. The first animation should show a contour plot of the $p(\mathbf{w}|\mathbf{t})$ posterior distribution as N increases. Each frame should indicate the true value $\mathbf{w}_{TRUE} = [1 \quad 1]^T$ (a vector of 1's; that is the value that was used to create the actual data set) by a $\circ$, the Maximum Likelihood estimate $\hat{\mathbf{w}}_{ML}$ as a $+$ and the MAP estimate $\hat{\mathbf{w}}_{MAP}$ as a $\times$. Each frame should include a legend for these two quantities. Choose 3-4 frames, include them in your report and comment extensively on them. In particular, what behaviors do you see and why are they expected?

(f) **[5 points]** Using the same sequential procedure, create a second MATLAB® animation that depicts the (x_n, t_n) data samples in the $x - t$ plane along with the associated 90% Prediction Interval (PI) as N increases from 0 to 1000. Once again, choose 3-4 frames, include them in your report and comment extensively on them.

Regarding the animations:

- For both animations, make sure the ranges of your x and y axis do not change from one frame to another; instead, find a good range and fix it.
- Save them in AVI format. Instead of uploading them to Canvas, upload them to another external location (*e.g.*, your public Dropbox, Google Drive, etc. folder) and provide a link that can be used to download them.

Task 2. [35 total points]
If several models are considered for a given regression problem, the most sensible way to select the best out of them is to use a separate hold-out set, called *validation* set, and compare them in terms of the sample Mean Squared Error (MSE) MSE_{val}, which is defined as the Sum of Squared Errors (SSE) computed using the validation samples and, finally, divided by the number of validation samples. Contrast this quantity to SSE_{train} divided by the number of training samples, denoted as MSE_{train}, which uses the training data to be computed and, in essence, is the loss function to minimize with respect to the parameters $\mathbf{w}$. The model selected as the best is the one with the smallest MSE_{val}. For whatever follows, consider again the `task1` data set.

(a) **[10 points]** Define the feature vector $\boldsymbol{\phi}_p(x) \triangleq \left[1 \; x \; x^2 \; x^3 \ldots x^p\right]^T$ for $p \geq 0$ (for $p = 0$, $\boldsymbol{\phi}_p(x) = 1$ for all x). Use Linear Regression (LR) on the first 10 samples of the data set using $\boldsymbol{\phi}_p(x)$ as the model's input data for $p = 0, \ldots, 12$. For a few values of p, show the graph of $y(x|\hat{\mathbf{w}}_{MLE})$ (model's output using the obtained MLE estimate of $\mathbf{w}$) versus x superimposed on the training data (add a legend to explain what is depicted). Comment on the obtained results.

(b) **[10 points]** Designate the remaining 990 samples in the `task1` data set as your validation set. Now, produce a single graph of MSE_{train} and the MSE_{val} of the previous LR models versus the number of model parameters $D = p + 1$. Amply comment on the results depicted in the graph. Why do you obtain these results?

(c) **[15 points]** Now consider a single-output RR model with regularization parameter $\mu \geq 0$ for the same data set with feature vector $\boldsymbol{\phi}_5(x)$. Again, produce a single graph of MSE_{train} and the MSE_{val} of the previous LR models versus μ. Notice that, for each value of μ, you will have to recompute the optimal weight vector $\mathbf{w}$ to compute both sample MSE values. Comment on (i) the behavior of the two graphs, (ii) the value μ^* of μ, which optimizes MSE_{val} and (iii) the optimal weight vector obtained using μ^* compared to $\mathbf{w}_{TRUE}$. How is RR justified?

Task 3. [30 total points]

This task pertains to the *Boston Housing* data set available through the UCI Machine Learning Repository (UCIMLR) [1]. In summary, the data set contains 506 samples, each having 13 attributes. The goal is to predict median values of owner-occupied homes in suburbs of Boston (MEDV column). Next, designate the first 306 samples as your training set, the next 100 as your validation set and the remaining 100 as your test set.

(a) **[10 points]** Consider a RR model with regularization parameter $\mu \geq 0$ that uses the original inputs as model inputs. Train the model for different values of μ, show a graph of MSE_{val} versus μ and identify the optimal value μ^* of μ as the one minimizing MSE_{val}. The model using the weights corresponding to μ^* will be the ones of your champion RR model.

(b) **[10 points]** Consider an RBF ANN that employs Gaussian interpolation kernels; therefore, the n^{th} model input is $\phi(\mathbf{x}, \mathbf{x}_n) = \exp\left\{-\gamma \left\|\mathbf{x} - \mathbf{x}_n\right\|_2^2\right\}$ for $n = 1, \ldots, N$, where $\gamma \geq 0$. Once again, train the model for different values of γ, show a graph of MSE_{val} versus γ and identify the optimal value γ^* of γ as the one minimizing MSE_{val}. The model using the weights corresponding to γ^* will be the ones of your champion RBF model.

(c) **[10 points]** Finally, compare the two champion models in terms of MSE_{val} and select a final winner. Evaluate on the test set and report the sample MSE value MSE_{test} of this model. How does it compare to the MSE_{test} of a LR model with $\mathbf{w} = \frac{1}{D}\mathbf{1}_D$, where $D = 13$? By the way, what does the latter model do? So, is the winning model you found useful in predicting housing values?

3 Preparation Guidelines

Below are some general guidelines that should be followed, when compiling a Mini-Project report. I strongly encourage you to stick to them, so that you receive full credit for your correct responses.

- **Task Statements:** Before attempting to address a particular task, ensure that you completely understand what is asked from you to perform and/or to produce. When in doubt, come to ask me 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 task/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 if you have to go back to it in the near future. Regarding the data you generate, keep them organize and document somehow (*e.g.*, in a text file) the specifics of how they were generated. **Caution:** You are not allowed to use any code and/or data that you have not produced without my explicit prior permission, in which case the sources you have obtained these from must be clearly indicated in your code or data description as well in your report. You are deemed to be plagiarizing, if you fail to do so, which may have dire consequences to your academic tenure here at Florida Tech!

- **Figures, Plots & Tables:** Plots should have their axes labeled and, if featuring several graphs, an appropriate legend should be used. Whether figures, plots 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 about what is being depicted? Apart from a concise description, major, relevant conclusions stemming from the display should also be included in the caption text.

- **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.

First, store all your Mini-Project deliverables in a folder named **lastname_mpX**, where "lastname" should be your last name and X should be the number of the Mini-Project, like 1, 2, etc. The folder name should be all lower case. For example, my folder for Mini-Project 1 would be named *anagnostopoulos_mp1*. Secondly, your **lastname_mpX** folder should have the following contents:

- 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 keep in mind:

– The report <u>must</u> include a signed & dated copy of the Work Origination Certification page. You can either scan such a page and include it in your document, or sign and date it electronically, as long as your signature is not typed. If this page is missing from your report, or it does not comply with the aforementioned conditions, I reserve the right not to accept the report and assign a score of 0/100 for the relevant Mini-Project.

– The Mini-Project may ask you to produce a variety of derivations, proofs, etc. You are not obliged to type such parts; it would be nice, but I realize that such effort would be quite time-consuming. Instead, you can import scanned images (or whole pages) of your handwritten work, as long as they are legible and well organized, so that the report has a clear logical flow. For example, it has to be clear where this hand-written work corresponds to (*e.g.*, which assignment it addresses).

– Having said all this, you may want to consider to print out your typed work, appropriately merge it with any handwritten pages (don't forget the signed Work Origination Certification page!) and then scan the whole compilation into a single PDF report, say, in the Library. **Caution:** when scanning, 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.

- A folder named **src**, which should contain all your MATLAB scripts that you authored and used for producing your results and the data sets that you created for this Mini-Project, if applicable.

- 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 Mini-Project report.

Next, compress your **lastname_mpX** folder into a single ZIP archive named **lastname_mpX.zip**; *e.g.*, mine would be called *anagnostopoulos_mp1.zip*.

Finally, upload your ZIP archive to Canvas by the specified deadline using the appropriate drop box. You are done!

References

[1] D. Harrison and D.L. Rubinfeld. Boston housing data, 1993. Accessed: 2017-3-16. URL: `http://www.dcc.fc.up.pt/~ltorgo/Regression/DataSets.html`.

[2] Mathworks. Animation, 2014. Accessed: 2017-3-16. URL: `http://www.mathworks.de/de/help/matlab/creating_plots/animation.html`.

WORK ORIGINATION CERTIFICATION

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

1. I have reviewed and understood the Academic Honesty section of the current version of FITs Student Handbook available at http://www.fit.edu/studenthandbook/, which discusses academic dishonesty (plagiarism, cheating, miscellaneous misconduct, etc.)
2. The content of this Mini-Project report 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 courses instructor.
3. In preparing and compiling all this report material, I have not collaborated with anyone and I have not received any type of help from anyone but from the courses instructor.

Signature ____ Date __________________