# AM41AN Coursework

Due by 25 April 2024 at 12:00 (electronic submission only)

## Instructions

In the present coursework you are required to perform a number of tasks. All your findings must be presented in the form of a scientific report. Such a report must be written in $\LaTeX$, and it must contain an abstract, introduction, methods, results and conclusions sections. All formulas must be typed (not scanned) and all your graphs and plots must be original. You are expected to work individually towards the completion of these tasks.

Each section must agree with the following descriptors:

- Abstract: It must summarize the nature of the problem, the methods used to solve it, the results obtained and the conclusions drawn from your results. **5 marks**.
- Introduction: It must contain a mathematical definition of the problem to be solved. Do not be afraid to use mathematics to properly define the problem (in this case, regression). Do not scan your formulas from somewhere else, type them yourself. Do not describe the particular instance you are solving (in the present case your data set has two-dimensional inputs, this is a particular feature of the present instance, the generic regression problem has inputs $\boldsymbol{x}$ in the space $\mathscr{X}$ which is a subset of $\mathbb{R}^d$, where $d$ is undetermined). **20 marks**.
- Methods: Thorough mathematical derivation and description of the method you will apply to solve the problem. You must describe your methods using mathematics. Word descriptions should appear, if so, after the math has done the talking. **20 marks**.
- Results: You present here your findings in the form of plots and tables. Axes must be named and both figures and tables must be captioned. Failure to do so negatively reflects in the presentation component of the assessment. **30 marks**.
- Conclusions: They must be extracted from the results presented in the section above. Conclusions are not a matter of opinion (“I think this could be seen as...”) they are matter of fact (“The results show ...”). **15 marks**.

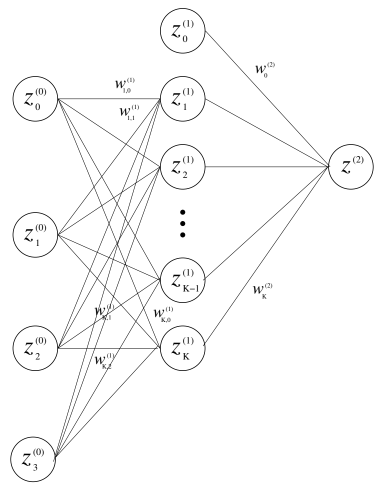

Figure 1: Network with four inputs, $K$ units in the hidden layer and one output.

- Presentation: Maximum of **10 marks**. Deviation from the rules presented above will result in mark deductions. Please do not forget to put your name and ID number at the top of your report.

The description of your tasks is as follows:

1. Multilayer Networks and EBP

   (a) Construct a data set $\mathcal{D}$ composed of $N = 50$ points of the form $(x_n, t_n)$ where $t_n = x_n + \frac{1}{4}\cos(\pi x_n) + \epsilon_n$, $\epsilon_n$ is drawn from the Gaussian distribution with zero mean and 0.1 variance, and $x_n$ is drawn from the uniform distribution in the domain $\mathscr{X} = (-1, 1)$. Present your results in the form of a scatter plot.

   (b) Consider the network presented in Figure 1 . The inputs correspond to the powers of $x$, i.e. $z_i^{(0)} = x^i$, $i = 0, 1, 2, 3$, the first unit in the hidden layer is set to 1, $z_0^{(1)} = 1$, the remaining $K$ hidden units compute the sigmoidal function

   $$z_k^{(1)} = \frac{1}{1 + \exp\left(-\sum_{j=0}^{3} w_{k,j}^{(1)} x^j\right)} \qquad k = 1, \ldots, K,$$

   and the output of the network is linear

   $$z^{(2)} = \sum_{k=0}^{K} w_k^{(2)} z_k^{(1)}.$$

   Compute the derivatives needed to implement the Error Back Propagation (EBP) algorithm supposing a sum-of-squares cost function $E = \frac{1}{2}\sum_{n=1}^{N}(t_n - z^{(2)}(x_n))^2$.

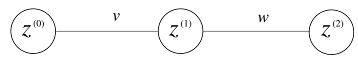


Figure 2: Feed-forward network with one input, one hidden unit and one output.

(c) Implement the EBP algorithm (feel free to modify appropriately the program provided for the Lab Session 7). The result of the training process should be presented in the form of curve $y : \mathscr{X} \to \mathbb{R}$ representing the expected behavior of $t$ given $x$. Consider the cases with $K = 3$, 5, and 11 units. Comment on your results.

(d) Give a justification for the use of the sum-of-square cost function (Hint: observe what type of noise have been used in the construction of the artificial data).

2. Regularization:

(a) Consider the network presented in Figure 2. The network is formed by an input $z^{(0)} = x$, a hidden unit $z^{(1)} = vz^{(0)}$ and an output $z^{(2)} = wz^{(1)}$. Given the cost function $E = \frac{1}{2}\sum_{n=1}^{N}(t_n - z^{(2)}(x_n))^2$ , compute the Hessian

$$\mathbf{H} = \left.\begin{pmatrix} \frac{\partial^2 E}{\partial v^2} & \frac{\partial^2 E}{\partial w \partial v} \\ \frac{\partial^2 E}{\partial v \partial w} & \frac{\partial^2 E}{\partial w^2} \end{pmatrix}\right|_{v,w=v_0,w_0},$$

where $w_0$ is the weight that produces the minimum cost. Consider the cases where the arithmetica average $\frac{1}{N}\sum_{n=1}^{N} x_n t_n$ is either positive or negative.

(b) Is the Hessian a positive definite matrix? What can you conclude from this observation?

3. Radial Basis Functions

(a) For the dataset constructed in Exercise 1(a) implement the Radial Basis Interpolation considering 5 centers on a regular grid. You may use (and modify) the programs presented in Lab Session 10. Present your result in the form of a curve $f : (-1, 1) \to \mathbb{R}$.

(b) Compare this curve with the one obtained by EBP (Exercise 1).