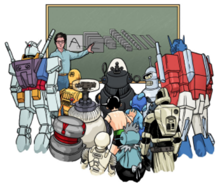
METU EE 449
Computational Intelligence



Training Artificial Neural Networks

Ozgur Gulsuna

[a] Middle East Technical University, Electrical and Electronics Engineering, Ankara, Turkey

---

Introduction

Insert here your abstract text.

Keywords: Type your keywords here, separated by semicolons ;

---

1. Basic Concepts

Here introduce the paper, and put a nomenclature if necessary, in a box with the same font size as the rest of the paper. The paragraphs continue from here and are only separated by headings, subheadings, images and formulae. The section headings are arranged by numbers, bold and 10 pt. Here follows further instructions for authors.

1.1. Which Function ?

An ANNs classifier that is trained with cross-entropy loss approximates the conditional probability distribution function. More specifically, for an input data, the output of the classifier is a probability distribution for the classes. The cross-entorpy loss function is a measure between the predicted probability distribution and the true distribution. The form of the loss function is decreasing, smooth and differentiable, which makes it easier to optimize using gradient-based methods. This form is also known as the negative log-like function.

1.2. Gradient Computation

1.3. Some Training Parameters and Basic Parameter Calculations

1. The batch refers to a subset of the training data that is used to compute the weights for one iteration. More specifically, the batch size is the number of training samples in a batch. The epoch on the other hand refers to the number of times the entire training data is used to update the weights. In training, there are generally multiple epoch iterations where the weights are updated with different batches/subsets of the training data.
2. For the N number of training samples, the number of batches per epoch is N/B, where B is the batch size. A little side note that the solution is rounded up to the higher integer if N/B is not an integer.
3. For the loss calculation iterations, such as SGD, for E number of epochs, the total number of iterations is $E \times N/B$. Again, a practical side note states that the N/B is rounded up to the higher integer.

E-mail address: ozgur.gulsuna@metu.edu.tr

1.4. Computing Number of Parameters of ANN Classifiers

1. Starting from the initial layer of the MLP, we have D_{in} number of input neurons and H_1 number of neurons in first hidden layer. Also there are biases associated with each neuron. Therefore, the number of parameters of the each layer is,

$$\begin{aligned}
\text{Input Layer} &= D_{in} \times H_1 \ + \ H_1 \\
\text{Hidden Layers} &= H_1 \times H_2 + H_2 \\
&\dots \\
\text{More Hidden Layers} &= H_{k-1} \times H_k + H_k \\
\text{Output Layer} &= H_k \times D_{out} \ + \ D_{out}
\end{aligned}$$

   The total sum can be written as, where K is the number of hidden layers.

$$\text{Total Number of Parameters} = D_{in} \times H_1 + \sum_{k=2}^{K} (H_{k-1} \times H_k + H_k) + H_k \times D_{out} + D_{out}$$

2. CNN structure is more complicated. The number of parameters of a CNN layer is calculated as follows:
   For the input layer, the number of parameters is,

$$\text{Input Layer} = (H_{in} \times W_{in} \times C_{in} \times C_1) + C_1$$

   where H_{in} and W_{in} are the height and width of the input image, and C_{in} is the number of channels of the input image. Each input of layer is the output of the previous layer. For the convolutional layers, the number of parameters is calculated as,

$$\text{Convolutional Layer} = H_k \times W_k \times C_{k-1} \times C_k + C_k$$

   Combination of all layers is,

$$\text{Convolutional Layers} = \sum_{k=2}^{K} H_k \times W_k \times C_{k-1} \times C_k + C_k$$

   Here all the parameters are summed up. The output is assumed to be the last index of the array. The final equation for the total number of parameters is,

$$\text{Total Number of Parameters} = (H_{in} \times W_{in} \times C_{in} \times C_1) + C_1 + \sum_{k=1}^{K} H_k \times W_k \times C_{k-1} \times C_k + C_k$$

2. Implementing a Convolutional Layer with NumPy

The section involves implementing conv2d function using NumPy for forward propagation and testing it on a small batch of MNIST dataset. We downloaded and loaded input and kernel files, and created an output image using the part2Plots function. The implementation code can be found in the appendix named my_conv2d.py. We confirmed the correctness of our implementation by the output image.

2.1. Section headings

Section headings should be left justified, bold, with the first letter capitalized and numbered consecutively, starting with the Introduction. Sub-section headings should be in capital and lower-case italic letters, numbered 1.1, 1.2, etc, and left justified, with second and subsequent lines indented. All headings should have a minimum of two text lines after them before a page or column break. Ensure the text area is not blank except for the last page.

2.2. General guidelines for the preparation of your text

Avoid hyphenation at the end of a line. Symbols denoting vectors and matrices should be indicated in bold type. Scalar variable names should normally be expressed using italics. Weights and measures should be expressed in SI units. All non-standard abbreviations or symbols must be defined when first mentioned, or a glossary provided.

2.3. File naming and delivery

Please title your files in this order 'procedia acronym_conference acronym_authorslastname'. Submit both the source file and the PDF to the Guest Editor.

Artwork filenames should comply with the syntax "aabbbbbb.ccc", where:

- a = artwork component type
- b = manuscript reference code
- c = standard file extension
  Component types:
- gr = figure
- pl = plate
- sc = scheme
- fx = fixed graphic

2.4. Footnotes

Footnotes should be avoided if possible. Necessary footnotes should be denoted in the text by consecutive superscript letters[1]. The footnotes should be typed single spaced, and in smaller type size (8 pt), at the foot of the page in which they are mentioned, and separated from the main text by a one line space extending at the foot of the column. The 'Els-footnote' style is available in the "TeX Template" for the text of the footnote.

Please do not change the margins of the template as this can result in the footnote falling outside printing range.

3. Illustrations

All figures should be numbered with Arabic numerals (1,2,3,....). Every figure should have a caption. All photographs, schemas, graphs and diagrams are to be referred to as figures. Line drawings should be good quality scans or true electronic output. Low-quality scans are not acceptable. Figures must be embedded into the text and not supplied separately. In MS word input the figures must be properly coded. Preferred format of figures are PNG, JPEG, GIF etc. Lettering and symbols should be clearly defined either in the caption or in a legend provided as part of the figure. Figures should be placed at the top or bottom of a page wherever possible, as close as possible to the first reference to them in the paper. Please ensure that all the figures are of 300 DPI resolutions as this will facilitate good output.

The figure number and caption should be typed below the illustration in 8 pt and left justified [***Note:*** one-line captions of length less than column width (or full typesetting width or oblong) centered]. For more guidelines and information to help you submit high quality artwork please visit: http://www.elsevier.com/artworkinstructions Artwork has no text along the side of it in the main body of the text. However, if two images fit next to each other, these may be placed next to each other to save space. For example, see Fig. 1.

[1] Footnote text.

Fig. 1. (a) first picture; (b) second picture.

4. Equations

Equations and formulae should be typed in MathType, and numbered consecutively with Arabic numerals in parentheses on the right hand side of the page (if referred to explicitly in the text). They should also be separated from the surrounding text by one space

$$\begin{aligned} X_r &= \dot{Q}''_{rad} \Big/ \left(\dot{Q}''_{rad} + \dot{Q}''_{conv} \right) \\ \rho &= \frac{\vec{E}}{J_c(T = \text{const.}) \cdot \left(P \cdot \left(\frac{\vec{E}}{E_c} \right)^m + (1 - P) \right)} \end{aligned} \tag{1}$$

5. Online license transfer

All authors are required to complete the Procedia exclusive license transfer agreement before the article can be published, which they can do online. This transfer agreement enables Elsevier to protect the copyrighted material for the authors, but does not relinquish the authors' proprietary rights. The copyright transfer covers the exclusive rights to reproduce and distribute the article, including reprints, photographic reproductions, microfilm or any other reproductions of similar nature and translations. Authors are responsible for obtaining from the copyright holder, the permission to reproduce any figures for which copyright exists.

Acknowledgements

Acknowledgements and Reference heading should be left justified, bold, with the first letter capitalized but have no numbers. Text below continues as normal.

Appendix A. An example appendix

Authors including an appendix section should do so before References section. Multiple appendices should all have headings in the style used above. They will automatically be ordered A, B, C etc.

A.1. Example of a sub-heading within an appendix

There is also the option to include a subheading within the Appendix if you wish.

References

[1] Filippini, Massimo, and Lester C. Hunt. (2011) "Energy demand and energy efficiency in the OECD countries: a stochastic demand frontier approach." *Energy Journal* **32** (2): 59–80.

[2] Filippini, Massimo, and Lester C. Hunt. (2012) "US residential energy demand and energy efficiency: A stochastic demand frontier approach." *Energy Economics* **34** (5): 1484–1491.

[3] Weyman-Jones, Thomas, Júlia Mendonça Boucinha, and Catarina Feteira Inácio. (2015) "Measuring electric energy efficiency in Portuguese households: a tool for energy policy." *Management of Environmental Quality: An International Journal* **26** (3): 407–422.

[4] Saunders, Harry (2009) "Theoretical Foundations of the Rebound Effect", in Joanne Evans and Lester Hunt (eds) *International Handbook on the Economics of Energy*, Cheltenham, Edward Elgar

[5] Sorrell, Steve (2009) "The Rebound Effect: definition and estimation", in Joanne Evans and Lester Hunt (eds) *International Handbook on the Economics of Energy*, Cheltenham, Edward Elgar

Instructions to Authors for LaTeX template:

1. ZIP mode for LaTeX template:

The zip package is created as per the guide lines present on the URL http://www.elsevier.com/author-schemas/preparing-crc-journal-articles-with-latex for creating the LaTeX zip file of Procedia LaTeX template. The zip generally contains the following files:

- ecrc.sty
- elsarticle.cls
- elsdoc.pdf
- .bst file
- Manuscript templates for use with these bibliographic styles
- Generic and journal specific logos, etc.

The LaTeX package is the main LaTeX template. All LaTeX support files are required for LaTeX pdf generation from the LaTeX template package.

Reference style .bst file used for collaboration support: In the LaTeX template packages of all Procedia titles a new ".bst" file is used which supports collaborations downloaded from the path http://www.elsevier.com/author-schemas/the-elsarticle-latex-document-class

2. Reference style used in Computer Science:

Title	Reference style
PROCS	3 Vancouver Numbered