Formatting Instructions for TMLR Journal Submissions

Anonymous authors
Paper under double-blind review

Abstract

The abstract paragraph should be indented 1/2 inch on both left and right-hand margins. Use 10 point type, with a vertical spacing of 11 points. The word **Abstract** must be centered, in bold, and in point size 12. Two line spaces precede the abstract. The abstract must be limited to one paragraph.

1 Submission of papers to TMLR

TMLR requires electronic submissions, processed by `https://openreview.net/`. See TMLR's website for more instructions.

If your paper is ultimately accepted, use option `accepted` with the `tmlr` package to adjust the format to the camera ready requirements, as follows:

`\usepackage[accepted]{tmlr}`.

You also need to specify the month and year by defining variables `month` and `year`, which respectively should be a 2-digit and 4-digit number. To de-anonymize and remove mentions to TMLR (for example for posting to preprint servers), use the preprint option, as in `\usepackage[preprint]{tmlr}`.

Please read carefully the instructions below, and follow them faithfully.

1.1 Style

Papers to be submitted to TMLR must be prepared according to the instructions presented here.

Authors are required to use the TMLR LaTeX style files obtainable at the TMLR website. Please make sure you use the current files and not previous versions. Tweaking the style files may be grounds for rejection.

1.2 Retrieval of style files

The style files for TMLR and other journal information are available online on the TMLR website. The file `tmlr.pdf` contains these instructions and illustrates the various formatting requirements your TMLR paper must satisfy. Submissions must be made using LaTeX and the style files `tmlr.sty` and `tmlr.bst` (to be used with LaTeX2e). The file `tmlr.tex` may be used as a "shell" for writing your paper. All you have to do is replace the author, title, abstract, and text of the paper with your own.

The formatting instructions contained in these style files are summarized in sections 2, 3, and 4 below.

2 General formatting instructions

The text must be confined within a rectangle 6.5 inches wide and 9 inches long. The left margin is 1 inch. Use 10 point type with a vertical spacing of 11 points. Computer Modern Bright is the preferred typeface throughout. Paragraphs are separated by 1/2 line space, with no indentation.

Paper title is 17 point, in bold and left-aligned. All pages should start at 1 inch from the top of the page.

Authors' names are set in boldface. Each name is placed above its corresponding address and has its corresponding email contact on the same line, in italic and right aligned. The lead author's name is to be listed first, and the co-authors' names are set to follow vertically.

Please pay special attention to the instructions in section 4 regarding figures, tables, acknowledgments, and references.

3 Headings: first level

First level headings are in bold, flush left and in point size 12. One line space before the first level heading and 1/2 line space after the first level heading.

3.1 Headings: second level

Second level headings are in bold, flush left and in point size 10. One line space before the second level heading and 1/2 line space after the second level heading.

3.1.1 Headings: third level

Third level headings are in bold, flush left and in point size 10. One line space before the third level heading and 1/2 line space after the third level heading.

4 Citations, figures, tables, references

These instructions apply to everyone, regardless of the formatter being used.

4.1 Citations within the text

Citations within the text should be based on the `natbib` package and include the authors' last names and year (with the "et al." construct for more than two authors). When the authors or the publication are included in the sentence, the citation should not be in parenthesis, using `\citet{}` (as in "See **?** for more information."). Otherwise, the citation should be in parenthesis using `\citep{}` (as in "Deep learning shows promise to make progress towards AI (**?**).").

The corresponding references are to be listed in alphabetical order of authors, in the **References** section. As to the format of the references themselves, any style is acceptable as long as it is used consistently.

4.2 Footnotes

Indicate footnotes with a number[1] in the text. Place the footnotes at the bottom of the page on which they appear. Precede the footnote with a horizontal rule of 2 inches.[2]

4.3 Figures

All artwork must be neat, clean, and legible. Lines should be dark enough for purposes of reproduction; art work should not be hand-drawn. The figure number and caption always appear after the figure. Place one line space before the figure caption, and one line space after the figure. The figure caption is lower case (except for first word and proper nouns); figures are numbered consecutively.

Make sure the figure caption does not get separated from the figure. Leave sufficient space to avoid splitting the figure and figure caption.

[1] Sample of the first footnote
[2] Sample of the second footnote

Table 1: Sample table title

PART	**DESCRIPTION**
Dendrite	Input terminal
Axon	Output terminal
Soma	Cell body (contains cell nucleus)

You may use color figures. However, it is best for the figure captions and the paper body to make sense if the paper is printed either in black/white or in color.

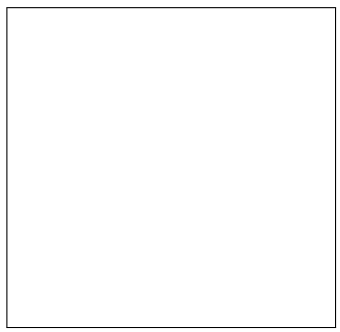


Figure 1: Sample figure caption.

4.4 Tables

All tables must be centered, neat, clean and legible. Do not use hand-drawn tables. The table number and title always appear before the table. See Table 1. Place one line space before the table title, one line space after the table title, and one line space after the table. The table title must be lower case (except for first word and proper nouns); tables are numbered consecutively.

5 Default Notation

In an attempt to encourage standardized notation, we have included the notation file from the textbook, *Deep Learning* **?** available at `https://github.com/goodfeli/dlbook_notation/`. Use of this style is not required and can be disabled by commenting out `math_commands.tex`.

Numbers and Arrays

a	A scalar (integer or real)
$\boldsymbol{a}$	A vector
$\boldsymbol{A}$	A matrix
$\mathsf{\mathbf{A}}$	A tensor
$\boldsymbol{I}_n$	Identity matrix with n rows and n columns
$\boldsymbol{I}$	Identity matrix with dimensionality implied by context
$\boldsymbol{e}^{(i)}$	Standard basis vector $[0, \dots, 0, 1, 0, \dots, 0]$ with a 1 at position i
$\text{diag}(\boldsymbol{a})$	A square, diagonal matrix with diagonal entries given by $\boldsymbol{a}$
a	A scalar random variable
$\mathbf{a}$	A vector-valued random variable
$\mathbf{A}$	A matrix-valued random variable

Sets and Graphs

$\mathbb{A}$	A set
$\mathbb{R}$	The set of real numbers
$\{0, 1\}$	The set containing 0 and 1
$\{0, 1, \dots, n\}$	The set of all integers between 0 and n
$[a, b]$	The real interval including a and b
$(a, b]$	The real interval excluding a but including b
$\mathbb{A} \backslash \mathbb{B}$	Set subtraction, i.e., the set containing the elements of $\mathbb{A}$ that are not in $\mathbb{B}$
$\mathcal{G}$	A graph
$Pa_{\mathcal{G}}(\mathrm{x}_i)$	The parents of x_i in $\mathcal{G}$

Indexing

a_i	Element i of vector $\boldsymbol{a}$, with indexing starting at 1
a_{-i}	All elements of vector $\boldsymbol{a}$ except for element i
$A_{i,j}$	Element i, j of matrix $\boldsymbol{A}$
$\boldsymbol{A}_{i,:}$	Row i of matrix $\boldsymbol{A}$
$\boldsymbol{A}_{:,i}$	Column i of matrix $\boldsymbol{A}$
$\mathsf{A}_{i,j,k}$	Element (i, j, k) of a 3-D tensor $\mathsf{\mathbf{A}}$
$\mathsf{\mathbf{A}}_{:,:,i}$	2-D slice of a 3-D tensor
a_i	Element i of the random vector $\mathbf{a}$

Calculus

$\dfrac{dy}{dx}$	Derivative of y with respect to x
$\dfrac{\partial y}{\partial x}$	Partial derivative of y with respect to x
$\nabla_{\boldsymbol{x}} y$	Gradient of y with respect to $\boldsymbol{x}$
$\nabla_{\boldsymbol{X}} y$	Matrix derivatives of y with respect to $\boldsymbol{X}$
$\nabla_{\mathsf{X}} y$	Tensor containing derivatives of y with respect to X
$\dfrac{\partial f}{\partial \boldsymbol{x}}$	Jacobian matrix $\boldsymbol{J} \in \mathbb{R}^{m\times n}$ of $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$
$\nabla_{\boldsymbol{x}}^2 f(\boldsymbol{x})$ or $\boldsymbol{H}(f)(\boldsymbol{x})$	The Hessian matrix of f at input point $\boldsymbol{x}$
$\int f(\boldsymbol{x}) d\boldsymbol{x}$	Definite integral over the entire domain of $\boldsymbol{x}$
$\int_{\mathbb{S}} f(\boldsymbol{x}) d\boldsymbol{x}$	Definite integral with respect to $\boldsymbol{x}$ over the set $\mathbb{S}$

Probability and Information Theory

$P(\mathrm{a})$	A probability distribution over a discrete variable
$p(\mathrm{a})$	A probability distribution over a continuous variable, or over a variable whose type has not been specified
$\mathrm{a} \sim P$	Random variable a has distribution P
$\mathbb{E}_{\mathrm{x}\sim P}[f(x)]$ or $\mathbb{E} f(x)$	Expectation of $f(x)$ with respect to $P(\mathrm{x})$
$\mathrm{Var}(f(x))$	Variance of $f(x)$ under $P(\mathrm{x})$
$\mathrm{Cov}(f(x), g(x))$	Covariance of $f(x)$ and $g(x)$ under $P(\mathrm{x})$
$H(\mathrm{x})$	Shannon entropy of the random variable x
$D_{\mathrm{KL}}(P \Vert Q)$	Kullback-Leibler divergence of P and Q
$\mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$	Gaussian distribution over $\boldsymbol{x}$ with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$

Functions

$f : \mathbb{A} \rightarrow \mathbb{B}$	The function f with domain $\mathbb{A}$ and range $\mathbb{B}$
$f \circ g$	Composition of the functions f and g
$f(\boldsymbol{x}; \boldsymbol{\theta})$	A function of $\boldsymbol{x}$ parametrized by $\boldsymbol{\theta}$. (Sometimes we write $f(\boldsymbol{x})$ and omit the argument $\boldsymbol{\theta}$ to lighten notation)
$\log x$	Natural logarithm of x
$\sigma(x)$	Logistic sigmoid, $\dfrac{1}{1+\exp(-x)}$
$\zeta(x)$	Softplus, $\log(1+\exp(x))$
$\Vert\boldsymbol{x}\Vert_p$	L^p norm of $\boldsymbol{x}$
$\Vert\boldsymbol{x}\Vert$	L^2 norm of $\boldsymbol{x}$
x^+	Positive part of x, i.e., $\max(0,x)$
$\boldsymbol{1}_{\mathrm{condition}}$	is 1 if the condition is true, 0 otherwise

6 Final instructions

Do not change any aspects of the formatting parameters in the style files. In particular, do not modify the width or length of the rectangle the text should fit into, and do not change font sizes (except perhaps in the **References** section; see below). Please note that pages should be numbered.

7 Preparing PostScript or PDF files

Please prepare PostScript or PDF files with paper size "US Letter", and not, for example, "A4". The -t letter option on dvips will produce US Letter files.

Consider directly generating PDF files using `pdflatex` (especially if you are a MiKTeX user). PDF figures must be substituted for EPS figures, however.

Otherwise, please generate your PostScript and PDF files with the following commands:

```
dvips mypaper.dvi -t letter -Ppdf -G0 -o mypaper.ps
ps2pdf mypaper.ps mypaper.pdf
```

7.1 Margins in LaTeX

Most of the margin problems come from figures positioned by hand using `\special` or other commands. We suggest using the command `\includegraphics` from the graphicx package. Always specify the figure width as a multiple of the line width as in the example below using .eps graphics

```
\usepackage[dvips]{graphicx} ...
\includegraphics[width=0.8\linewidth]{myfile.eps}
```

or

```
\usepackage[pdftex]{graphicx} ...
\includegraphics[width=0.8\linewidth]{myfile.pdf}
```

for .pdf graphics. See section 4.4 in the graphics bundle documentation (`http://www.ctan.org/tex-archive/macros/latex/required/graphics/grfguide.ps`)

A number of width problems arise when LaTeX cannot properly hyphenate a line. Please give LaTeX hyphenation hints using the `\-` command.

Broader Impact Statement

In this optional section, TMLR encourages authors to discuss possible repercussions of their work, notably any potential negative impact that a user of this research should be aware of. Authors should consult the TMLR Ethics Guidelines available on the TMLR website for guidance on how to approach this subject.

Author Contributions

If you'd like to, you may include a section for author contributions as is done in many journals. This is optional and at the discretion of the authors. Only add this information once your submission is accepted and deanonymized.

Acknowledgments

Use unnumbered third level headings for the acknowledgments. All acknowledgments, including those to funding agencies, go at the end of the paper. Only add this information once your submission is accepted and deanonymized.

8 Simulations

8.1 Linear regression posterior estimation with synthetic data

The purpose of these simulations is to test that VIFA is able to learn the posterior distribution of a very simple linear regression model with two learnable parameters.

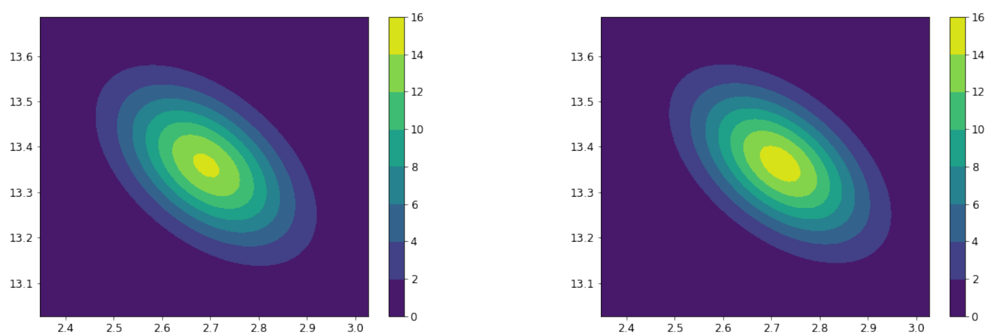
8.1.1 Methodology

Synthetic data was generated as follows. First, 1000 inputs $\boldsymbol{x} \in \mathbb{R}^2$ were sampled from a multivariate zero mean Gaussian distribution with unit variances and covariances of 0.5. Next, the linear regression parameter vector $\boldsymbol{\theta} \in \mathbb{R}^2$ was sampled from $\mathcal{N}(\boldsymbol{\theta}; 0, \alpha^{-1}\boldsymbol{I})$ with $\alpha = 0.01$. Then the outputs $y \in \mathbb{R}$ were generated according to the equation $y = \boldsymbol{\theta}^T\boldsymbol{x} + \epsilon$, where $\epsilon \sim \mathcal{N}(\epsilon; 0, \beta^{-1})$ with $\beta = 0.1$.

Using this data, the true posterior distribution was evaluated in closed form (TODO: needs citation or reference to equation) and an approximate posterior with latent dimension $K = 1$ was estimated via VIFA. VIFA ran for 5000 epochs with a batch size of $M = 100$ and a Monte Carlo average size of $L = 10$. The learning rates $\eta_{\boldsymbol{c}}$, $\eta_{\boldsymbol{F}}$ and $\eta_{\boldsymbol{\gamma}}$ were set to 0.01, 0.0001 and 0.01, respectively. The reasoning for using a smaller learning rate for $\boldsymbol{F}$ was that its contribution to the full covariance matrix is $\boldsymbol{F}\boldsymbol{F}^T$. Since this is regression, the likelihood function used in VIFA was set to $\mathcal{N}(y; \boldsymbol{\theta}^T\boldsymbol{x}, \beta^{-1})$. Finally, to improve numerical stability, any gradients with Frobenius norm greater than 10 were rescaled to have norm of exactly 10.

8.1.2 Results and discussion

Figure 2 shows a qualitative comparison between the ground truth and approximate linear regression posteriors. (TODO: add discussion here plus results for more random seeds in appendix).



(a) Ground truth

(b) VIFA

Figure 2: The ground truth posterior pdf of a linear regression model with two learnable parameters, plus the pdf of a FA model with a single latent dimension which was fit to the same data using VIFA.

8.2 Logistic regression posterior estimation with synthetic data

The purpose of these simulations is to test that VIFA is able to learn the posterior distribution of a very simple logistic regression model with two learnable parameters.

8.2.1 Methodology

Synthetic inputs $\boldsymbol{x} \in \mathbb{R}^2$ and the logistic regression parameter vector $\boldsymbol{\theta} \in \mathbb{R}^2$ were sampled in the same way as in Section 8.1.1. However, this time 3000 inputs were used. To generate each output $y \in \{0, 1\}$, the positive class probability was first evaluated as $p = \sigma(\boldsymbol{\theta}^T \boldsymbol{x})$, where $\sigma : \mathbb{R} \to [0, 1]$ denotes the logistic sigmoid function. Then a random number was sampled uniformly from the interval $[0, 1]$ and y was set to 1 if this number was less than p, else 0.

Unlike linear regression, there is no closed form solution for the true posterior of a logistic regression model. In this case, the ground truth posterior was evaluated by first looping over a 2D grid around the true parameter vector $\boldsymbol{\theta}$ and evaluating the unnormalised log posterior probability at each point in the grid. Formally, this is

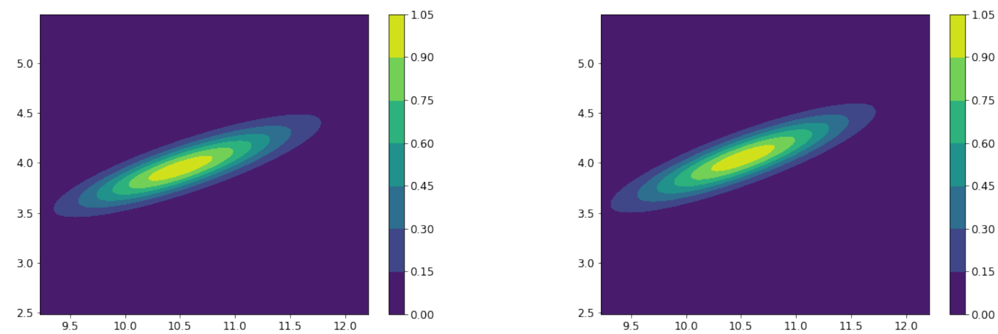
$$\mathcal{N}(\boldsymbol{\theta}; 0, \alpha^{-1}\boldsymbol{I}) \prod_{n=1}^{3000} \sigma(\boldsymbol{\theta}^T \boldsymbol{x}_n)^{y_n} \sigma(\boldsymbol{\theta}^T \boldsymbol{x}_n)^{1-y_n}. \tag{1}$$

Then the values in the grid were scaled such that the maximum value was equal to 1. This posterior is only correct up to a constant, but suffices for a qualitative comparison.

The posterior was then approximated via VIFA using the exact same hyperparameters as in Section 8.1.1. This time, the likelihood function used in VIFA was set to binary cross-entropy. That is, $\sigma(\boldsymbol{\theta}^T \boldsymbol{x}_n)^{y_n} \sigma(\boldsymbol{\theta}^T \boldsymbol{x}_n)^{1-y_n}$.

8.2.2 Results and discussion

Figure 3 shows a qualitative comparison between the ground truth and approximate logistic regression posteriors. (TODO: add discussion here plus results for more random seeds in appendix).



(a) Ground truth

(b) VIFA

Figure 3: The ground truth posterior pdf of a logistic regression model with two learnable parameters, plus the pdf of a FA model with a single latent dimension which was fit to the same data using VIFA. Both posteriors are scaled such that the maximum value is equal to 1.

A Appendix

You may include other additional sections here.