Project 9: Structural Health Monitoring

This manuscript (<u>permalink</u>) was automatically generated from <u>dbudzik/project9 SHM@b12c719</u> on December 4, 2020.

Authors

- David Budzik
 - · 🕝 dbudzik
- Claudia Mederos
 - · 🕝 <u>cr25</u>
- Bolaji Lawal
 - O bolajilawal
- Abdullah Assaf
 - aboo12

Machine Learning Methods

4. Methodology

4.1 Explanation of the Problem:

Our Kaggle Project was focused on Binary classification. Binary Classification is the task of classifying the elements of a set into two groups on the basis of a classification rule. Our problem revolved around classifying the structural damage detected in a structure. We had to choose different machine learning algorithms to help classify the data.

4.2 Steps For each algorithm:

First we had to create a model, softmax or sigmoid used in output layer, so that we could train the model and tune hyperparameters, such as accuracy, precision etc. Finally, we had to test the performance of model against the sample data. This allowed us to compare results from different models and see what model was the most accurate at classifying the data.

4.3 Data Preparation:

First we had to split the train dataset into 2, one set for training and the other for validation. Then we had to shuffle the train dataset before using it. We then used regularizations to apply penalties on layer parameters or layer activity during optimization. Dropouts were used to incorporate non-linearity. The data was normalized by subtracting the mean and dividing by standard deviation.

4.4 Method 1: Logistic Regression:

The first Method utilized was Logistic Regression. It was the simplest model for binary classification. The method outputs a probability between 0 and 1, exclusivly. It then defines a decision threshold to map to binary category. The logistic regression yielded great results on the first run with an accuracy of \sim 96%, low precision \sim 24%, and recall plummeting to \sim 20%.

4.5 Method 2: Artificial Neural Networks:

The second method utilized was Artificial Neural Networks. Artificial Neural Networks are more complex than logistic regression. This method adds a bunch of hidden non-linear layers to the logistic regression model. Our group used this method to check if it offered an improvement on the previous model. The artificial neural networks yeiled average results with great accuracy ~96%, Improved precision ~45%, Improved recall ~60%, although the metrics were not good overall.

4.6 Method 3: Random Forest Regression:

The third method utilized was Random Forest Regression. Random Forest Regression performs both regression and classification tasks with the use of multiple decision trees and bagging. It is easy to use and often returns good results even without hyperparameter tuning. Our group used this method to check if it offered an improvement on the previous model. The Random Forest Regression provided the worst results with extremely inaccurate accuracy rate, while also not working well for the type of data we had.

4.7 Issue:

We noticed that the problem was the precision and recall are very low even though accuracy is high. This was caused by the damaged data being too sparse, only 3.5% of our data represents the damaged condition. The solution was too add copies of the damaged data into the training set.

4.8 Take Two:

In take two we ran logistic regression and artifical nueral networks again to see if it yeilded better results with addition of new copies of damaged data. Logistic Regression returned good accuracy ~83%, much better precision ~60%, and Fantastic recall ~99%. Artifical Neural Network returned slighly better data with a fantastic accuracy ~98%, fantastic precision ~91%, and fantastic recall ~99%.

Results

Results go here

Discussion of Results

Discussion goes here

Conclusion

Conclusion goes here

References

1. Sci-Hub provides access to nearly all scholarly literature

Daniel S Himmelstein, Ariel Rodriguez Romero, Jacob G Levernier, Thomas Anthony Munro, Stephen Reid McLaughlin, Bastian Greshake Tzovaras, Casey S Greene

eLife (2018-03-01) https://doi.org/ckcj

DOI: 10.7554/elife.32822 · PMID: 29424689 · PMCID: PMC5832410

2. Reproducibility of computational workflows is automated using continuous analysis

Brett K Beaulieu-Jones, Casey S Greene

Nature biotechnology (2017-04) https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6103790/

DOI: <u>10.1038/nbt.3780</u> · PMID: <u>28288103</u> · PMCID: <u>PMC6103790</u>

3. Bitcoin for the biological literature.

Douglas Heaven

Nature (2019-02) https://www.ncbi.nlm.nih.gov/pubmed/30718888

DOI: 10.1038/d41586-019-00447-9 · PMID: 30718888

4. Plan S: Accelerating the transition to full and immediate Open Access to scientific publications

cOAlition S

(2018-09-04) https://www.wikidata.org/wiki/Q56458321

5. Open access

Peter Suber *MIT Press* (2012)

ISBN: <u>9780262517638</u>

6. Open collaborative writing with Manubot

Daniel S. Himmelstein, Vincent Rubinetti, David R. Slochower, Dongbo Hu, Venkat S. Malladi, Casey S. Greene, Anthony Gitter

Manubot (2020-05-25) https://greenelab.github.io/meta-review/

7. Opportunities and obstacles for deep learning in biology and medicine

Travers Ching, Daniel S. Himmelstein, Brett K. Beaulieu-Jones, Alexandr A. Kalinin, Brian T. Do, Gregory P. Way, Enrico Ferrero, Paul-Michael Agapow, Michael Zietz, Michael M. Hoffman, ... Casey S. Greene

Journal of The Royal Society Interface (2018-04-04) https://doi.org/gddkhn

DOI: <u>10.1098/rsif.2017.0387</u> · PMID: <u>29618526</u> · PMCID: <u>PMC5938574</u>

8. Open collaborative writing with Manubot

Daniel S. Himmelstein, Vincent Rubinetti, David R. Slochower, Dongbo Hu, Venkat S. Malladi, Casey S. Greene, Anthony Gitter

PLOS Computational Biology (2019-06-24) https://doi.org/c7np

DOI: 10.1371/journal.pcbi.1007128 · PMID: 31233491 · PMCID: PMC6611653

This manuscript is a template (aka "rootstock") for <u>Manubot</u>, a tool for writing scholarly manuscripts. Use this template as a starting point for your manuscript.

The rest of this document is a full list of formatting elements/features supported by Manubot. Compare the input (.md files in the /content directory) to the output you see below.

Basic formatting

Bold text

Semi-bold text

Centered text

Right-aligned text

Italic text

Combined italics and bold

Strikethrough

- 1. Ordered list item
- 2. Ordered list item
 - a. Sub-item
 - b. Sub-item
 - i. Sub-sub-item
- 3. Ordered list item
 - a. Sub-item
- · List item
- · List item
- List item

subscript: H₂O is a liquid

superscript: 2¹⁰ is 1024.

unicode superscripts 0123456789

unicode subscripts 0123456789

A long paragraph of text. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Putting each sentence on its own line has numerous benefits with regard to <u>editing</u> and <u>version</u> <u>control</u>.

Line break without starting a new paragraph by putting two spaces at end of line.

Document organization

Document section headings:

Heading 1

Heading 2

Heading 3

Heading 4

Heading 5

Heading 6



Horizontal rule:

Heading 1's are recommended to be reserved for the title of the manuscript.

Heading 2's are recommended for broad sections such as Abstract, Methods, Conclusion, etc.

Heading 3's and Heading 4's are recommended for sub-sections.

Links

Bare URL link: https://manubot.org

<u>Long link with lots of words and stuff and junk and bleep and blah and stuff and other stuff and more stuff yeah</u>

Link with text

Link with hover text

Link by reference

Citations

Citation by DOI [1].

Citation by PubMed Central ID [2].

Citation by PubMed ID [3].

Citation by Wikidata ID [4].

Citation by ISBN [5].

Citation by URL [6].

Citation by alias [7].

Multiple citations can be put inside the same set of brackets [1,5,7]. Manubot plugins provide easier, more convenient visualization of and navigation between citations [2,3,7,8].

Citation tags (i.e. aliases) can be defined in their own paragraphs using Markdown's reference link syntax:

Referencing figures, tables, equations

Figure 1

Figure 2

```
Figure 3

Figure 4

Table 1

Equation 1

Equation 2
```

Quotes and code

Quoted text

Quoted block of text

Two roads diverged in a wood, and I—I took the one less traveled by, And that has made all the difference.

Code in the middle of normal text, aka inline code.

Code block with Python syntax highlighting:

```
from manubot.cite.doi import expand_short_doi

def test_expand_short_doi():
    doi = expand_short_doi("10/c3bp")
    # a string too long to fit within page:
    assert doi == "10.25313/2524-2695-2018-3-vliyanie-enhansera-copia-i-
        insulyatora-gypsy-na-sintez-ernk-modifikatsii-hromatina-i-
        svyazyvanie-insulyatornyh-belkov-vtransfetsirovannyh-geneticheskih-
        konstruktsiyah"
```

Code block with no syntax highlighting:

```
Exporting HTML manuscript
Exporting DOCX manuscript
Exporting PDF manuscript
```

Figures



Figure 1: A square image at actual size and with a bottom caption. Loaded from the latest version of image on GitHub.



Figure 2: An image too wide to fit within page at full size. Loaded from a specific (hashed) version of the image on GitHub.



Figure 3: A tall image with a specified height. Loaded from a specific (hashed) version of the image on GitHub.



Figure 4: A vector .svg image loaded from GitHub. The parameter sanitize=true is necessary to properly load SVGs hosted via GitHub URLs. White background specified to serve as a backdrop for transparent sections of the image.

Tables

Table 1: A table with a top caption and specified relative column widths.

Bowling Scores	Jane	John	Alice	Bob
Game 1	150	187	210	105
Game 2	98	202	197	102
Game 3	123	180	238	134

Table 2: A table too wide to fit within page.

	Digits 1-33	Digits 34-66	Digits 67-99	Ref.
р	i 3.14159265358979323 846264338327950	28841971693993751 0582097494459230	78164062862089986 2803482534211706	piday.org
е	2.71828182845904523 536028747135266	24977572470936999 5957496696762772	40766303535475945 7138217852516642	nasa.gov

 Table 3: A table with merged cells using the attributes plugin.

	Colors	
Size	Text Color	Background Color
big	blue	orange
small	black	white

Equations

A LaTeX equation:

$$\int_0^\infty e^{-x^2} dx = \frac{\sqrt{\pi}}{2} \tag{1}$$

An equation too long to fit within page:

$$x = a + b + c + d + e + f + g + h + i + j + k + l + m + n + o + p + q + r + s + t + u + v + w + x + y + z + 1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9$$
(2)

Special

▲ WARNING The following features are only supported and intended for .html and .pdf exports. Journals are not likely to support them, and they may not display correctly when converted to other formats such as .docx.

LINK STYLED AS A BUTTON

Adding arbitrary HTML attributes to an element using Pandoc's attribute syntax:

Manubot Manubot Manubot Manubot Manubot. Manubot Manubot Manubot Manubot. Manubot Manubot Manubot. Manubot Manubot. Manubot.

Adding arbitrary HTML attributes to an element with the Manubot attributes plugin (more flexible than Pandoc's method in terms of which elements you can add attributes to):

Manubot Manubot.

Available background colors for text, images, code, banners, etc:

white lightgrey grey darkgrey black lightred lightyellow lightgreen lightblue lightpurple red orange yellow green blue purple

Using the Font Awesome icon set:



Light Grey Banner useful for *general information* - <u>manubot.org</u>

1 Blue Banner

useful for *important information* - <u>manubot.org</u>

♦ Light Red Banner useful for *warnings* - <u>manubot.org</u>