Assessment 3: Project

Dominic Shore   
Canberra, Australia  
DominicShore@gmail.com

# Introduction

The task of assessing whether an abalone has had the opportunity to reach sexual maturity and reproduce can be accomplished by “cutting the shell through the cone, staining it, and counting the number of rings through a microscope”[1]. If such a model yielded a sufficiently high level of accuracy it could help regulatory authorities to calibrate minimum size requirements in order to maximise the value of sustainably harvested abalone subject to the constraint of reproduction.

This paper presents a series of models which were designed to predict the age range of abalone from physical measurements. All models were built with neural networks using Keras and sklearn libraries in python.

# methodology

## Data source, ingestion and transformation

The abalone data was sourced from the University of California Irvine online repository of datasets[2]. In order to ensure that the results were not unduly influenced by the presences of outliers in the data each variable was examined. Using a combination of summary statistics and data visualisation the variable ‘Height’ was found to have two obvious outliers. Those outliers were then removed from the data and subsequent calculations.

Chart

Description automatically generated

1. Outliers in the ‘Height’ variable are shown in the Figure above

The variable of ‘rings’ – denoting the age of the abalone – was converted into a categorical variable as follows by assigning each observation to one of four categories using the following approach:

* Class 1: 0 - 7 years
* Class 2: 8- 10 years
* Class 3: 11 - 15 years
* Class 4: Greater than 15 years

This new variable was named ‘ring\_class’ and will serve as the target variable for the predictions from the models in this paper. To help the models deal with a categorical target variable more easily it was converted using the OneHotEncoder method in sklearn. New variables were created for each category. The ring class variable one-hot encoding can be represented graphically as follows:

A picture containing table

Description automatically generated

1. The transformation of the ring class varaible with one-hot encoding

Similarly, the variable ‘sex’ was transformed using sklearn’s OrdinalEncoder to alter the representation of the flags M (for Male), F (Female), and I (Infant) to 0, 1, and 2 respectively.

Finally, the remaining numerical variables (length, diameter, height, whole weight, shucked weight, viscera weight, & shell weight) were normalised so as to range between the values of 0 and 1. This was achieved with the use of sklearn’s MinMaxScaler.

## Data exploration

The distribution of ‘ring\_class’ was found to be contain a large number of observations in Class 2 (8-10 years). This may be because of the previous minimum size restrictions which have historically been placed on the harvest[4] of abalone.

Chart, bar chart

Description automatically generated

1. Distribution of abalone ring count classes

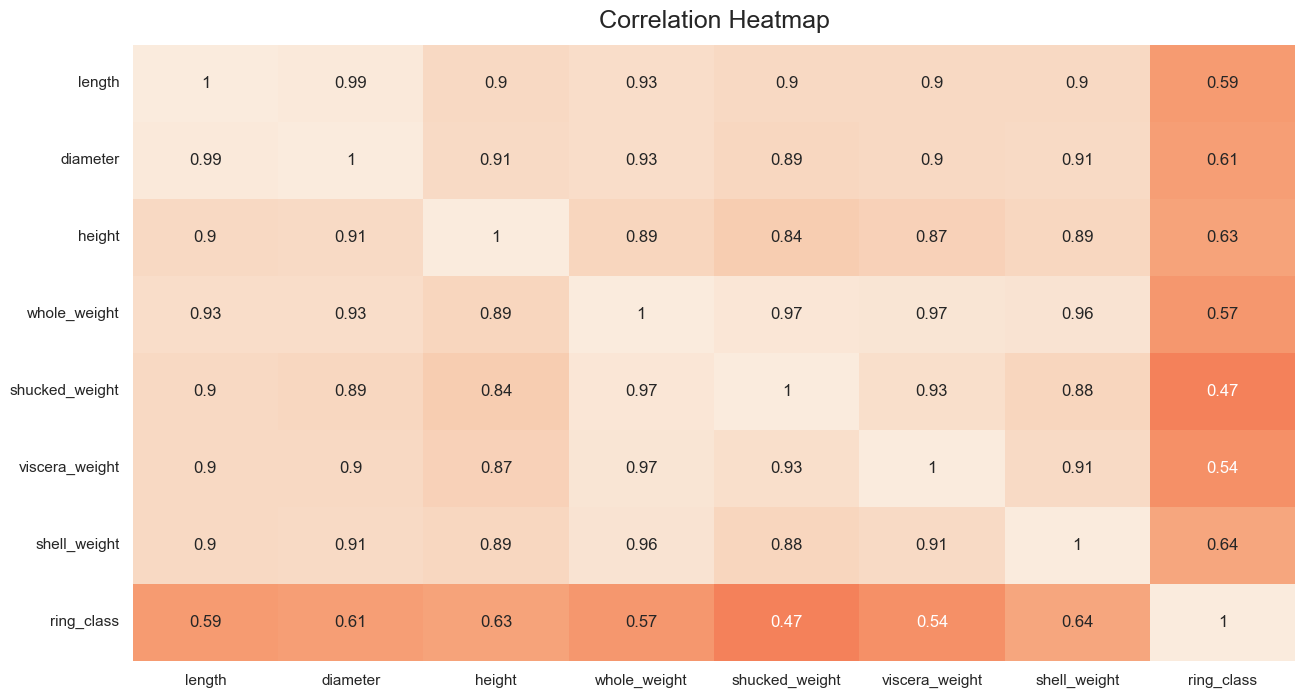
The distribution of each of the variables in the abalone dataset was then examined graphically using histograms:

Chart, bar chart, histogram

Description automatically generated

1. Histogram of each variable included in the abalone dataset from UCI

To understand the relationships between the variables inmore depth a correlation heatmap was developed with the aid of the seaborn library.



1. Correlation heatmap of the abalone dataset from UCI

The correlation heatmap of the abalone dataset showed that the shell height and shell weight were the most highly correlated with the ring class variable.

## Model contruction

Four related models were constructed to assess the impact of changes in i) hidden neurons, ii) learning rates, iii) hidden layers, and iv) optimisers on the accuracy of the model in predicting the correct ring class.

Each model had an 8-node input layer to accommodate ingestion of the 8 features related to the size, weight and sex of abalone in the dataset.

The initial round of experiments involved changes to the number of neurons in a single hidden layer. These changes were made in an iterative fashion to investigate the effect of differences in model capacity on accuracy. The testing included a hidden layer of size 5, 10, 15, and 20 with each being subject to a *relu* activation function before being passed to the output layer. In all cases, the output layer consisted of four neurons which corresponded to the ring class categories of the target variable.

The second round of experiments keeps the optimal number of hidden neurons observed during the initial round constant and makes changes to the learning rate hyperparameter to further optimise the accuracy of the model. The learning rates experimented with in this round were 1e-1, 1e-2, 1e-3, 1e-4, and finally 1e-5.

The third round of experiments investigated the impact of adding a second hidden layer to the network architecture while maintaining the optimal number of hidden layer neurons (from the initial experiment) and the optimal learning rate (from the pervious experiment).

Finally, the optimiser was changed in light of previously optimal architecture and hyperparameter settings. In the first three (i – iii) experiments on model accuracy, Stochastic Gradient Descent (SGD) was employed to optimise the weight and bias terms in the neural network. The SGD method entails one random instance of the data at a time which is passed forward through the network of neurons. Next the error resulting from that forward pass is calculated by comparing the actual and desired output from the network. Then, by using the chain rule to calculate the contribution of each output to the total error, the algorithm performs a reverse pass to modify each parameter by moving it by a small fraction in the opposite direction of the gradient calculated using that rule[3].

In this final step the Adaptive Moment Estimator (Adam) was applied to the model. “[Adam] … is perhaps best seen as a variant on the combination of RMSProp and momentum with a few important distinctions. First, in Adam, momentum is incorporated directly as an estimate of the ﬁrst-order moment (with exponential weighting) of the gradient. The most straightforward way to add momentum to RMSProp is to apply momentum to the rescaled gradients. The use of momentum in combination with rescaling does not have a clear theoretical motivation. Second, Adam includes bias corrections to the estimates of both the ﬁrst-order moments (the momentum term) and the (uncentered) second-order moments to account for their initialization at the origin”[4].

In the development of each model 10 experimental runs with different training datasets were performed. The mean and standard deviation of the accuracy metrics where calculated and noted. The hyperparameters from the most accurate model were employed in subsequent models as the outlined above were implemented.

# results

In

A close-up of a necklace

Description automatically generated with low confidence

1. Representation of the final single layer net with 20 hidden nodes

# Conclusion

Before you begin to format your paper, first write and save the content as a separate text file. Complete all content and organizational editing before formatting. Please note sections A-D below for more information on proofreading, spelling and grammar.

Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit use of hard returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads-the template will do that for you.

## Abbreviations and Acronyms

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

## Units

* Use either SI (MKS) or CGS as primary units. (SI units are encouraged.) English units may be used as secondary units (in parentheses). An exception would be the use of English units as identifiers in trade, such as “3.5-inch disk drive”.
* Avoid combining SI and CGS units, such as current in amperes and magnetic field in oersteds. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.
* Do not mix complete spellings and abbreviations of units: “Wb/m2” or “webers per square meter”, not “webers/m2”. Spell out units when they appear in text: “. . . a few henries”, not “. . . a few H”.
* Use a zero before decimal points: “0.25”, not “.25”. Use “cm3”, not “cc”. (*bullet list*)

## Equations

The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled.

Number equations consecutively. Equation numbers, within parentheses, are to position flush right, as in (1), using a right tab stop. To make your equations more compact, you may use the solidus ( / ), the exp function, or appropriate exponents. Italicize Roman symbols for quantities and variables, but not Greek symbols. Use a long dash rather than a hyphen for a minus sign. Punctuate equations with commas or periods when they are part of a sentence, as in:

*a**b* 

Note that the equation is centered using a center tab stop. Be sure that the symbols in your equation have been defined before or immediately following the equation. Use “(1)”, not “Eq. (1)” or “equation (1)”, except at the beginning of a sentence: “Equation (1) is . . .”

## Some Common Mistakes

* The word “data” is plural, not singular.
* The subscript for the permeability of vacuum **0, and other common scientific constants, is zero with subscript formatting, not a lowercase letter “o”.
* In American English, commas, semicolons, periods, question and exclamation marks are located within quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.)
* A graph within a graph is an “inset”, not an “insert”. The word alternatively is preferred to the word “alternately” (unless you really mean something that alternates).
* Do not use the word “essentially” to mean “approximately” or “effectively”.
* In your paper title, if the words “that uses” can accurately replace the word “using”, capitalize the “u”; if not, keep using lower-cased.
* Be aware of the different meanings of the homophones “affect” and “effect”, “complement” and “compliment”, “discreet” and “discrete”, “principal” and “principle”.
* Do not confuse “imply” and “infer”.
* The prefix “non” is not a word; it should be joined to the word it modifies, usually without a hyphen.
* There is no period after the “et” in the Latin abbreviation “et al.”.
* The abbreviation “i.e.” means “that is”, and the abbreviation “e.g.” means “for example”.

An excellent style manual for science writers is [7].

# Using the Template

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

## Authors and Affiliations

**The template is designed for, but not limited to, six authors.** A minimum of one author is required for all conference articles. Author names should be listed starting from left to right and then moving down to the next line. This is the author sequence that will be used in future citations and by indexing services. Names should not be listed in columns nor group by affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization).

### For papers with more than six authors: Add author names horizontally, moving to a third row if needed for more than 8 authors.

### For papers with less than six authors: To change the default, adjust the template as follows.

#### Selection: Highlight all author and affiliation lines.

#### Change number of columns: Select the Columns icon from the MS Word Standard toolbar and then select the correct number of columns from the selection palette.

#### Deletion: Delete the author and affiliation lines for the extra authors.

## Identify the Headings

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include Acknowledgments and References and, for these, the correct style to use is “Heading 5”. Use “figure caption” for your Figure captions, and “table head” for your table title. Run-in heads, such as “Abstract”, will require you to apply a style (in this case, italic) in addition to the style provided by the drop down menu to differentiate the head from the text.

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced. Styles named “Heading 1”, “Heading 2”, “Heading 3”, and “Heading 4” are prescribed.

## Figures and Tables

#### Positioning Figures and Tables: Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation “Fig. 1”, even at the beginning of a sentence.

1. Table Type Styles

| Table Head | Table Column Head | | |
| --- | --- | --- | --- |
| Table column subhead | Subhead | Subhead |
| copy | More table copya |  |  |

1. Sample of a Table footnote. (*Table footnote*)
2. Example of a figure caption. (*figure caption*)

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity “Magnetization”, or “Magnetization, M”, not just “M”. If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write “Magnetization (A/m)” or “Magnetization {A[m(1)]}”, not just “A/m”. Do not label axes with a ratio of quantities and units. For example, write “Temperature (K)”, not “Temperature/K”.

##### Acknowledgment *(Heading 5)*

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression “one of us (R. B. G.) thanks ...”. Instead, try “R. B. G. thanks...”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

##### References

The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use “Ref. [3]” or “reference [3]” except at the beginning of a sentence: “Reference [3] was the first ...”

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors’ names; do not use “et al.”. Papers that have not been published, even if they have been submitted for publication, should be cited as “unpublished” [4]. Papers that have been accepted for publication should be cited as “in press” [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

1. W. J. Nash, “The Population Biology of Abalone (Haliotis species) in Tasmania. I. Blacklip Abalone (H. rubra) from the North Coast and the Islands of Bass Strait” Department of Primary Industry And Fisheries, Technical Report No. 48. 1994
2. Dua, D. and Graff, C. (2019). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.
3. D. Rumelhart et al. (1985) “Learning Internal Represenational by Error Propagation,” Defense Tehnical Information Cneter technical report, September.
4. I. Goodfellow, Y. Bengio and A. Courville (2016) “Deep learning,” MIT Press [http://www.deeplearningbook.org] Chapter 8, pg.305
5. Fishing Tasmainia (accessed 07/10/2022), “Abalone Fishing” [https://fishing.tas.gov.au/recreational-fishing/fishing-by-species/abalone/abalone-fishing]
6. Yedla, Anurag & Davoudi Kakhki, Fatemeh & Jannesari, Ali. (2020). Predictive Modeling for Occupational Safety Outcomes and Days Away from Work Analysis in Mining Operations. International Journal of Environmental Research and Public Health. 17. 7054. 10.3390/ijerph17197054.
7. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
8. M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

**IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from your paper may result in your paper not being published.**