Abalone Classification

Ryan Heslin, Chenxi Liao, Sebastian Zovko

April 19, 2022

Introduction

The goal of this analysis is to create an inferential model of the sex of abalone, a type of saltwater mollusk. This is a classification problem with three classes: infant, (adult) male, and (adult) female. Avaliable features include:

- Measurements of abalone dimensions
- The weights of different parts of the abalone, as well as the entire animal
- The number of rings of the abalone (abalone grow rings as they age)

More on the Problem

Wild abalone populations have collapsed from overharvesting, but they are farmed commercially for their meat and pearls. Predicting sex is of interest because the eggs of adult female abalone are useful to breeders and scientists (Bradley, 2010).

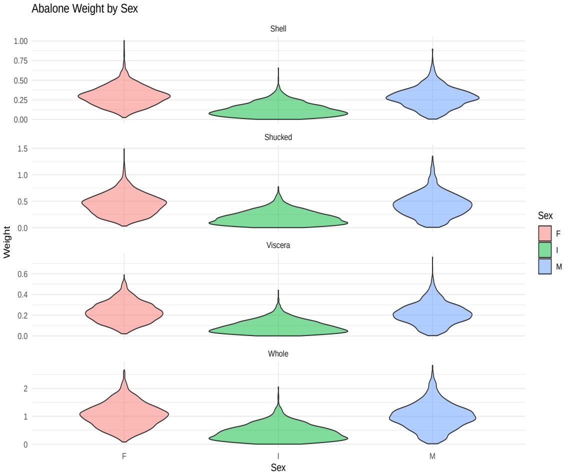
Available Predictors

- Length
- Diameter
- Height
- Weight
    - Whole
    - Viscera
    - Shucked
    - Shell
- Rings (roughly corresponds to abalone age)

Abalone Weight by Sex

Infant abalone weigh substantially less than adults, but male and female adult abalone weigh about the same. This illustrates the main challenge of this classification task: infant abalone are easy to distinguish from adults, but male and female adults are hard to distinguish from each other.

Abalone Weight by Sex Visualized



The Data in Detail

The full data have 4177 observations. The dataset (Kaggle, 2020) is standard in machine learning research. It may be obtained here. The three separate weight variables don't quite sum to total weight, suggesting measurement error. Still, including all three would be a bad idea because it would make the model matrix nearly singular. We will select one to use. This table summariezes the difference of whole weight and the sum of the other weight variables

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-0.447	0.018	0.037	0.05	0.068	0.608

General Approach

We fit an initial $J - 1$ (here $J = 3$) logits baseline model using sex, dimensions, and whole weight as predictors. This model is appropriate for nominal outcomes, which abalone sex is because it has no logical ordering. This model consists of $J - 1$ logit models, each predicting the log odds of a given class versus a baseline class. There are $\binom{J}{2}$ possible comparisons, but all can be derived algebraically from just $J - 1$ logits. We choose infants as the reference class because they differ in the same way from each adult class (i.e., are younger).

Initial Model Fitting and Data Partitioning

y.level	term	estimate	statistic	p.value
F	(Intercept)	-2.953	-5.362	0.000
F	Whole.weight	5.393	10.737	0.000
F	Length	-16.578	-5.279	0.000
F	Diameter	10.992	2.836	0.005
F	Height	8.402	2.258	0.024
F	Rings	0.228	8.502	0.000
M	(Intercept)	-0.238	-0.559	0.576
M	Whole.weight	6.314	13.449	0.000
M	Length	-17.915	-5.962	0.000
M	Diameter	6.361	1.703	0.089
M	Height	5.226	1.444	0.149
M	Rings	0.212	8.088	0.000

Overall Likelihood Ratio Test of Initial Model

The overall likelihood ratio test of the initial model is highly significant, so we reject the null hypothesis of no linear relationship

Model	Resid. df	Resid. Dev	Test	Df	LR stat.	Pr(Chi)
Null	6680	7325		NA	NA	NA
Full	6670	5764	1 vs 2	10	1561	0

Wald Tests of Coefficients

Most of the estimated coefficients are significant under Wald tests, even after applying the Bonferroni correction to *p*-values.

y.level	term	estimate	statistic	p.value
F	(Intercept)	-2.953	-5.36	0.000
F	Whole.weight	5.393	10.74	0.000
F	Length	-16.578	-5.28	0.000
F	Diameter	10.992	2.84	0.055
F	Height	8.402	2.26	0.287
F	Rings	0.228	8.50	0.000

Choosing a Model by Stepwise Selection

Stepwise selection by AIC chooses a model with a few interactions. The upper scope includes all pairwise interactions and quadratic variable terms.

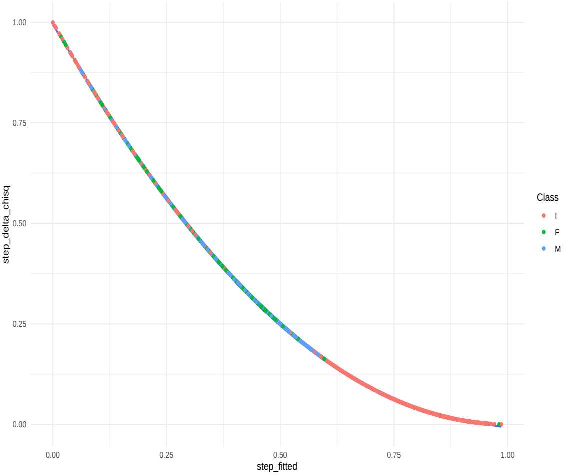
The Stepwise-Selected Model

y.level	term	estimate	statistic	p.value
F	(Intercept)	-5.740	-6.034	0.000
F	Whole.weight	8.333	4.537	0.000
F	Length	-20.821	-6.056	0.000
F	Diameter	3.752	0.898	0.369
F	Height	38.576	3.561	0.000
F	Rings	0.951	9.042	0.000

Residuals by Class

Average standardized residuals are notably lower for infants than for adults, consistent with infants being easier to classify.

Delta Chi-Square vs. Predicted Probability



Model Testing

A likelihood-ratio test of the step-selected model over the initial model is highly significant. A goodness-of-fit test for the step-selected model has a p-value that is computationally 1, providing no evidence against the null of a good fit

```
Statistic: 112
Degrees of freedom: 6
Threshhold: 12.6
p = 6.87e-22
```

Type II ANOVA

A type II ANOVA test shows that all predictors aside from `diameter` significantly improve the fit when included. `Whole.weight` is by far the most important, followed by `Rings` and the interactions.

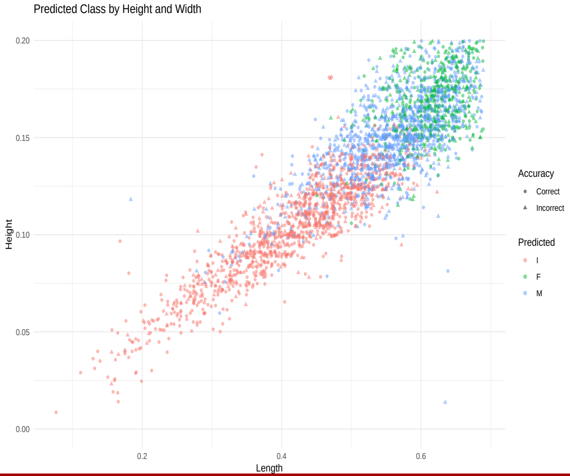
	LR Chisq	Df	Pr(>Chisq)
Whole.weight	288.53	2	0.000
Length	52.77	2	0.000
Diameter	4.48	2	0.106
Height	5.60	2	0.061
Rings	80.75	2	0.000
Whole.weight:Rings	16.05	2	0.000
Height:Rings	12.57	2	0.002
Whole.weight:Diameter	4.61	2	0.100

Comparing Different Weight Variables

Next we try refitting the model with each of all four available weight variables (including replacing the interaction) and comparing AIC. (AIC has no inherent interpretation, but is useful when comparing variants of the same model). It turns out whole weight has the lowest, but the differences are minor.

Class Separation by Height and Width

Plotting predicted class by height and width shows again that infants are well separated from adults. # Predicted Class by Height and Width



Comparison with Binary Classification

It is obvious that distinguising infants from adults is much easier than male from female adults. For comparison, we combine the male and female classes into "adult" and fit a binomial model.

term	estimate	std.error	p.value
(Intercept)	-2.787	0.713	0.000
Whole.weight	9.048	1.733	0.000
Length	-21.965	3.178	0.000
Diameter	1.013	3.835	0.792
Height	32.752	9.570	0.001
Rings	0.889	0.092	0.000
Whole.weight:Rings	-0.409	0.087	0.000
Height:Rings	-2.909	0.890	0.001
Whole.weight:Diameter	6.554	2.908	0.024

Deviance Comparison

While a direct likelihood ratio test is inappropriate because these models use different versions of the response, it is worth noting this binary model has less than half the deviance of the step-selected model.

Binary Deviance	Three-Class Deviance
2560	5652

Model Validation

On both training and testing sets, sensitivity, specificity, and precision are much higher for the infant than the adult classes. However, test error was only a little higher than train error. Still, overall test accuracy is above 50%, much better than the 36% (the highest class proportion) achieved by the naive classifier.

	I	F	M
I	859	48	162
F	154	330	566
M	240	288	694

Overall accuracy: 0.564

Class-Specific Statistics for Training Set

	I	F	M
sensitivity	0.804	0.314	0.568
specificity	0.830	0.683	0.692
precision	0.686	0.495	0.488

Test Set Results

Results on the test set are very similar, suggesting minimal generalization error. Overall test accuracy is 0.568.

	I	F	M
I	223	14	36
F	36	82	139
M	67	69	170

	I	F	M
sensitivity	0.817	0.319	0.556
specificity	0.834	0.692	0.692
precision	0.684	0.497	0.493

Test-Train Comparison and Overall Accuracy

The difference of the two confusion matrices shows that test and train performance are similar.

	I	F	M
sensitivity	-0.013	-0.005	0.012
specificity	-0.005	-0.009	0.001
precision	0.002	-0.001	-0.005

Conclusion

Male and female adult abalone are difficult to distinguish using predictors avaliable in this dataset. Using a $J-1$ logits baseline model with interaction terms, approximately 50% classification accuracy was achieved. However, infant abalone are smaller than adults of either sex, making them significantly easier to classify accurately. Future research should focus on finding easily measured predictors for which the adult sexes are easily distinguished. Published research tends to focus on classifying abalone age in years (e.g, Abdelbar [1998]). For the reason noted above, this in some ways an easier problem than classifying sex.

References

Abalone Dataset. (n.d.). [2022]. Kaggle. Retrieved April 15, 2022, from https://www.kaggle.com/rodolfomendes/abalone-dataset

Abdelbar, A. M. (1998). Achieving superior generalisation with a high order neural network. Neural Computing and Applications, 7, 141–147.

Bradley, R. (2010, March 9). How to Sex an Abalone: A Sea Snail's Story. The Atlantic. https://www.theatlantic.com/health/archive/2010/03/how-to-sex-an-abalone-a-sea-snails-story/37198/