

Lab 4 GSI

Names of group members *

Zihao Chen, Todd Faulkenberry, Yutong Wang

Readability of report (5 points) *

	1	2	3	4	5	
Narrative unclear and/or difficult to read	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	Narrative very clear and/or easy to read

Grammar of report (5 points) *

	1	2	3	4	5	
Incorrect written grammar pervasive	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	Excellent written grammar

EDA and model choices

Exploratory data analysis (4 points) *

	0	1	2	3	4	
Did not provide any exploratory figures or numerical summaries of the data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	Provided clear, relevant figures and summaries of the data

Comments about EDA

It would have been smart to only choose one of the images to visualize for the expert labels and the 3 features that way you could have done it all in one plot.

Other than that it looks good!

Justification of variable selection *

	0	1	2	3	
Provided no figures, justification or discussion of variable selection	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	Described clearly and thoughtfully which variables are best and provided insightful figures

Comments about variable selection

This discussion is thorough, but a bit wordy.

Also, the correlation table would have been nice if it had been done as sort of a heat map where you color each entry in the table by the strength of the correlation

Appropriateness of prediction methods *

	0	1	2	3	
Did not discuss appropriateness of methods chosen	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	Clearly outlined the assumptions and reasons for choosing each model

Comments on classifiers

You needed to discuss the assumptions of the models and whether they were violated in this problem (e.g. in logistic regression the independence assumption is likely violated)

Model performance

Depth of exploration concerning model fit and convergence *

	0	1	2	3	4	
Did not discuss model fit	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	Clearly described how well each model fit from a variety of different angles. Provided informative and high quality visualizations.

Though about how to appropriately use cross-validation *

	0	1	2	3	
Did not consider CV carefully	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	Clearly outlined that the pixels should be grouped in some way when doing CV

Comments on model fit and convergence

In the model fit part it would have been good to look at true positive, false positive, etc. or make an ROC curve or something else other than the CV accuracy

For CV I was a bit confused of what you decided to do in the end, but good job recognizing that you should block the data.

Depth of exploration on patterns in misclassification errors *

	0	1	2	3	4	
Did not explore patterns	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	Clearly explored and visualized patterns in misclassification errors

Comments on patterns in misclassification

It would have been good to see what was going on at the feature level in the misclassified points.

Justification of using model on future data *

	0	1	2	3	
Did not justify answer to whether or not the model would work well on future data	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	Clearly explained why or why not the model would work on future data

Comments on using model on future data

I had some trouble understand how you split the data but in the end it seems like you decided to do CV with all of the data so you should point out that you did not have a held out test set to see how you would do on future data

Reproducibility

Everything was provided in order for reproducibility *

	0	1	2	3	
Did not provide all files needed	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	Provided all files necessary and clearly labelled how to reproduce all analyses (i.e. which files produce what and how they all fit together)

Comments on reproducibility

Conclusion

One or more things that were done well

Good job with the EDA and explaining yourselves throughout the report

One or more things that could be improved upon

There were some places where you could be more concise in your explanations that would make the report easier to follow

Other comments

This form was created inside of UC Berkeley.

Google Forms