Assignment 1 ML
March 22, 2022

J L Gouws
19G4436

1. See script file for details.

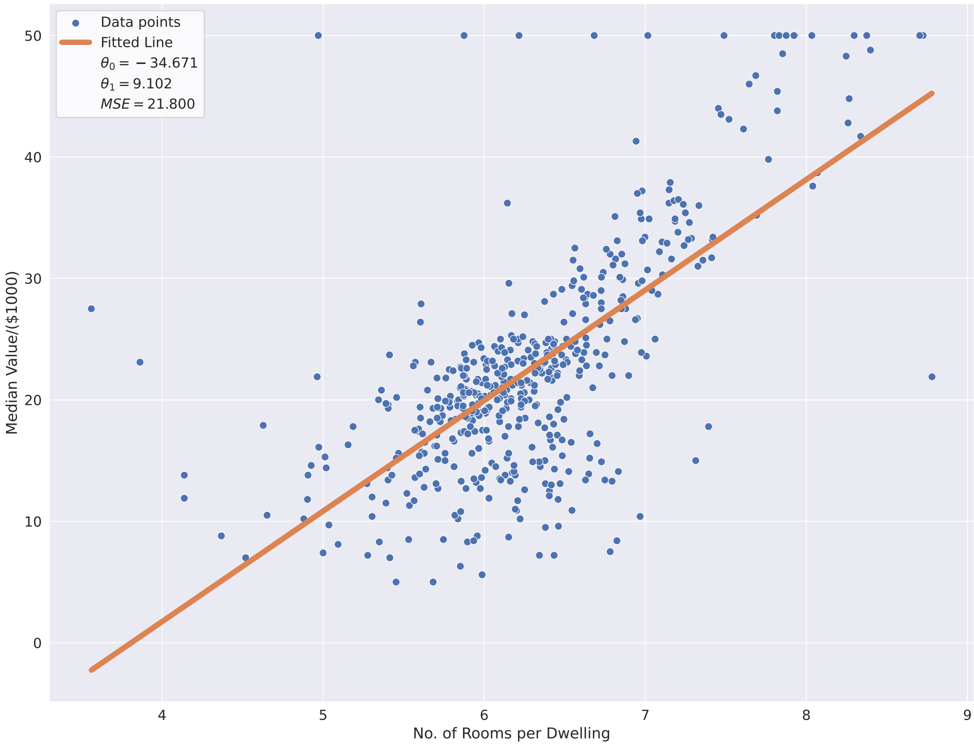

Figure 1: The relation between RM and MEDV with a fitted regression line.

2. a) See script file.

b) It is pretty much the same, the testing errors are the same to the first four significant figures.
This can be seen by looking at the MSE(on test data) in Table 3.

c) The plot of the two graphs is also more or less the same.
The plot for Lasso regression is significantly different, however.
We can see the values of the parametes of the two variables in Table 2. We can see the difference in the training methods for the Bivariate case in Figure 2.

d) The normal linear model seemed to train faster, but the Ridge model's test went faster. That being said, the difference in testing speed could be down to random error. There are other processes running on my computer which might interfere with accurate timing

e) Yes, the Lasso method has a significantly larger MSE, and the model also trained a lot slower.

f) We can see in Table 3, Table 5 and Table 6 that the MSE does change as Lasso and Ridge parameter changes. The generalization error actually increases as we change the parameters. The training and testing times do not change, however.

g) The mean absolute error:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|$$

This is very similar to the MSE. Only the errors are not squared. For MSE, squaring a large absolute error makes the error disproportionately large, which punishes an outliers. Since large errors are not squared for MAE, the outliers are punished but not over excessively.

h) The adjusted R^2 metric tries to deal with this problem, the formula is:

$$R^2_{\text{adj}} = 1 - \frac{(1-R^2)(N-1)}{N-p-1}$$

Where N is the total sample size, and p is the number of features of the model.

i) We cannot compare this to the results of Task 1. Task 1 is a bivariate linear regression and does not depend on B.

We can compare this to part a) for the multi-linear case. We can see the difference in the parameters when 'B' is dropped by comparing Table 2 and Table 4. We can see that the multidimensional plot changes for the multilinear case, MEDV depends differently on RM.

Table 1: Different Linear Regression Models

Model	MSE	R^2	Train Time $\times 10^{-6}$	Test Time $\times 10^{-6}$
Linear	22.737590	0.759814	368.595123	74.386597
Ridge	22.923754	0.763344	409.841537	68.187714
Lasso	27.606505	0.701360	460.863113	71.048737

Table 2: Parameters of Different Linear Regression Models

Model	θ_0	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}	θ_{11}	θ_{12}	θ_{13}
Linear	36.73	-0.12	0.04	0.02	2.78	-18.59	3.76	0.00	-1.47	0.31	-0.01	-0.95	0.01	-0.55
Ridge	31.47	-0.12	0.04	-0.01	2.57	-10.79	3.80	-0.00	-1.36	0.30	-0.01	-0.86	0.01	-0.56
Lasso	40.55	-0.08	0.04	-0.00	0.00	-0.00	1.01	0.02	-0.64	0.27	-0.01	-0.74	0.01	-0.78

Table 3: Different Linear Regression Models when B is Excluded

Model	MSE	R^2	Train Time $\times 10^{-6}$	Test Time $\times 10^{-6}$
Linear	23.425683	0.775628	464.200974	53.882599
Ridge	23.628806	0.777696	337.362289	51.975250
Lasso	28.433794	0.710581	428.676605	56.266785

Table 4: Parameters of Different Linear Regression Models With B Excluded

Model	θ_0	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}	θ_{11}	θ_{13}
Linear	42.31	-0.13	0.04	0.01	2.98	-19.42	3.59	0.01	-1.48	0.29	-0.01	-0.93	-0.58
Ridge	36.95	-0.13	0.04	-0.02	2.77	-11.28	3.64	0.00	-1.37	0.27	-0.01	-0.84	-0.59
Lasso	45.02	-0.09	0.04	-0.00	0.00	-0.00	0.88	0.03	-0.64	0.26	-0.02	-0.72	-0.81

Table 5: Different Linear Regression Models with $\alpha = 2$

Model	MSE	R^2	Train Time $\times 10^{-6}$	Test Time $\times 10^{-6}$
Linear	22.737590	0.759814	406.980515	68.664551
Ridge	23.108689	0.762833	329.732895	66.995621
Lasso	32.515507	0.639673	385.046005	69.856644

Table 6: Different Linear Regression Models with $\alpha = 3$

Model	MSE	R^2	Train Time $\times 10^{-6}$	Test Time $\times 10^{-6}$
Linear	22.737590	0.759814	388.145447	67.710876
Ridge	23.237427	0.761974	329.017639	66.518784
Lasso	34.971519	0.623029	356.912613	70.095062

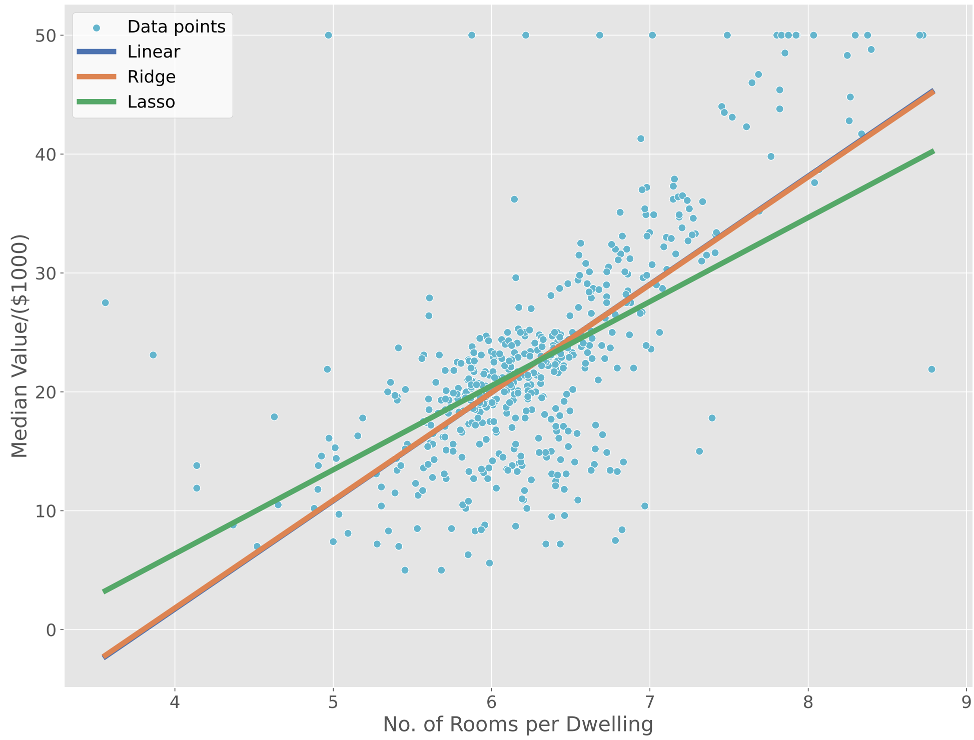

Figure 2: The relation between RM and MEDV for different regression methods.

3 and 4. A quick note on the results. The classifier for task 3 and task 4 has an accuracy of 100%. This makes sense because if we look at plots of features, the type of plant can be discerned by eye, so the computer will definitely be able to do it.

5. a) See Figure 3

b) We probably could use them to make a semi-fine classifier. There is not too much overlap between malignant and benign cases.

c) See Figure 4.

d) We probably could use these to make a semi-fine classifier aswell. There is not too much overlap between malignant and benign cases. That being said, medical things have to be pretty accurate, so I am not sure if these will suffice.

e) I chose area and perimeter, the data looked more separate for these two.
The F1 score seemed better for for these features than other features, when I trained models on other features.

f) See Figure 5 for the values.
Accuracy: 0.8882
Precision: 0.8363
Recall: 0.8614
F1: 0.8487

g) I would say recall. Recall measures how many of the cases are detected as malignant. In the case of something that is life threatening, it is much better to detect all the cases, even if some benign tumors are labeled as malignant.

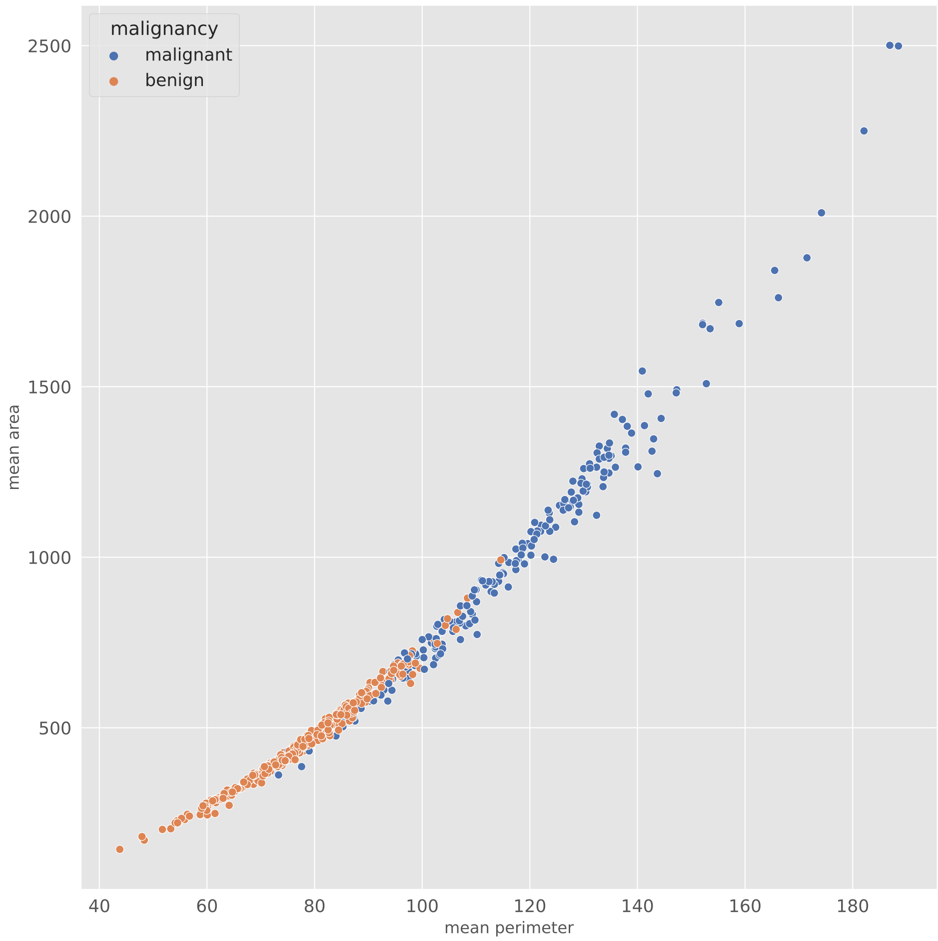

Figure 3: Perimeter Vs Area of Tumor

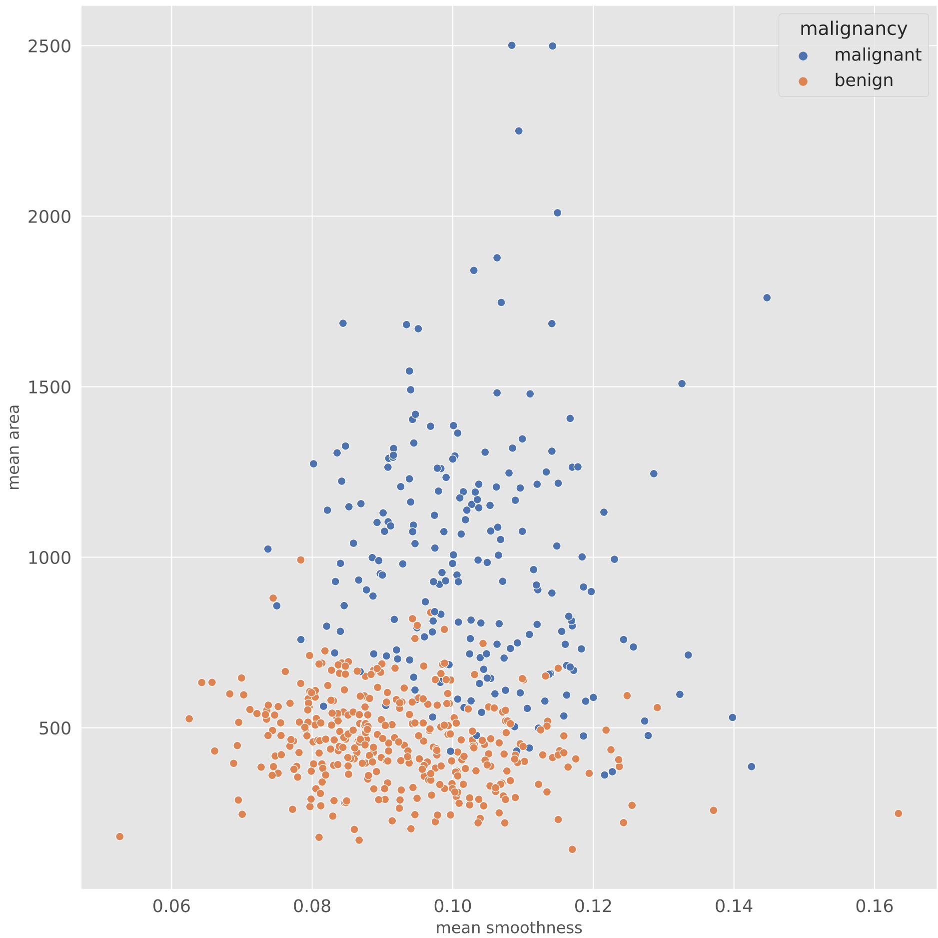

Figure 4: Smoothness Vs Area of Tumor

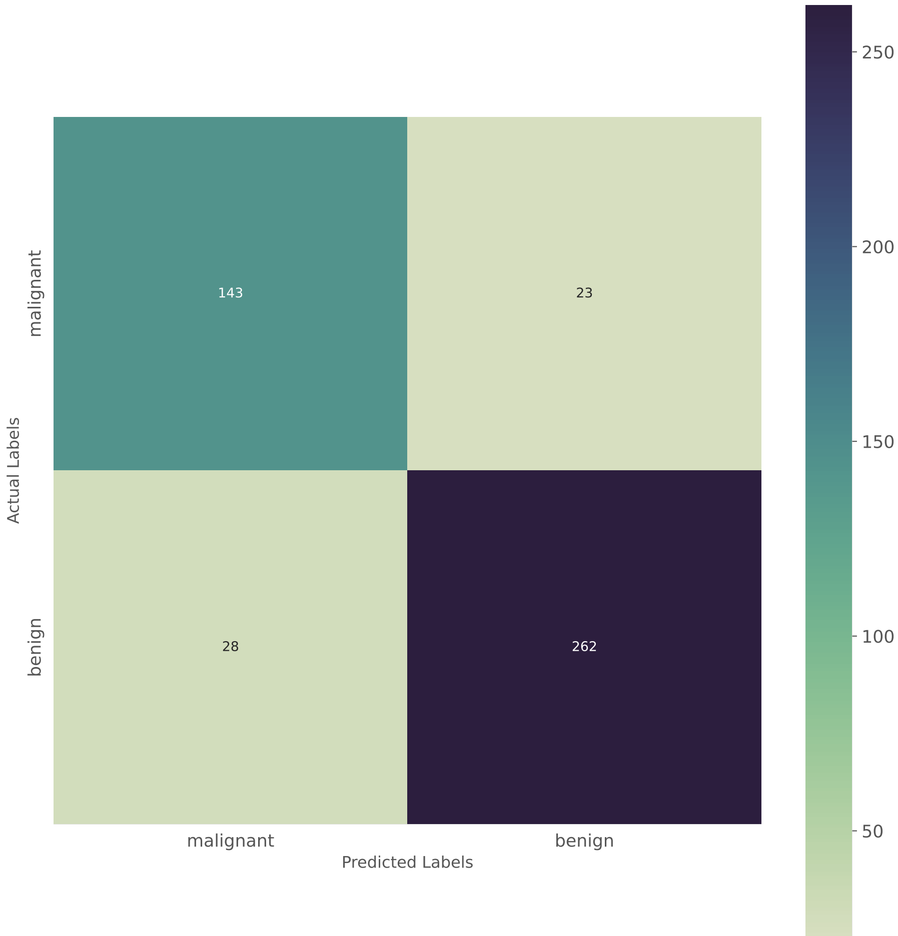

Figure 5: Confusion Matrix for the Breast Cancer classifier