HW2

Peter Buonaiuto

October 2023

1 Linear and Ridge Regression

1.a Fit a hyperplane on the data

$\widetilde{w} = (\widetilde{D}^T \widetilde{D})^{-1} \widetilde{D}^T Y$

$$\widetilde{D}^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 5 & 1 & 1 & 2 & 2 \\ 1 & 3 & 5 & 1 & 1 \\ 2 & 5 & 3 & 1 & 2 \end{bmatrix}$$

$$\widetilde{D} = \begin{bmatrix} 1 & 5 & 1 & 2 \\ 1 & 1 & 3 & 5 \\ 1 & 1 & 5 & 3 \\ 1 & 2 & 1 & 1 \\ 1 & 2 & 1 & 2 \end{bmatrix}$$

$$Y = \begin{bmatrix} 5 \\ 3 \\ 1 \\ 2 \\ 2 \end{bmatrix}$$

$$\widetilde{D}^T \cdot \widetilde{D} = \begin{bmatrix} 5 & 11 & 11 & 13 \\ 11 & 35 & 17 & 24 \\ 11 & 17 & 37 & 35 \\ 13 & 24 & 35 & 43 \end{bmatrix}$$

$$(\widetilde{D}^T \cdot \widetilde{D})^{-1} = \begin{bmatrix} 3.02 & -0.56 & -0.32 & -0.34 \\ -0.56 & 0.15 & 0.07 & 0.03 \\ -0.32 & 0.07 & 0.15 & -0.07 \\ -0.34 & 0.03 & -0.07 & 0.17 \end{bmatrix}$$

$$(\widetilde{D}^T \widetilde{D})^{-1} \cdot \widetilde{D}^T = \begin{bmatrix} -0.78 & -0.20 & -0.15 & -1.24 & 0.90 \\ 0.32 & -0.06 & 0.03 & -0.16 & -0.13 \\ 0.06 & -0.13 & 0.32 & -0.09 & -0.16 \\ 0.05 & 0.32 & -0.16 & -0.19 & -0.02 \end{bmatrix}$$

$$\widetilde{w} = (\widetilde{D}^T\widetilde{D})^{-1}\widetilde{D}^TY = \begin{bmatrix} -0.408 \\ 0.873 \\ -0.269 \\ 0.646 \end{bmatrix}$$

Bias b is $w_0 = -0.408$

The hyperplane is therefore represented by the fitted model
$Y = -0.408 + 0.873X_1 - 0.269X_2 + 0.646X_3$

1.b Ridge Regression

If we use ridge regression with $\alpha = 0.1$ instead, we must recalculate by adding α down the diagonal of $\widetilde{D}^T\widetilde{D}$.

$$\widetilde{D}^T\widetilde{D} + \alpha I = \begin{bmatrix} 5.1 & 11 & 11 & 13 \\ 11 & 35.1 & 17 & 24 \\ 11 & 17 & 37.1 & 35 \\ 13 & 24 & 35 & 43.1 \end{bmatrix}$$

$$(\widetilde{D}^T\widetilde{D} + \alpha I)^{-1} = \begin{bmatrix} 2.29 & -0.42 & -0.24 & -0.26 \\ -0.42 & 0.13 & 0.06 & 0.01 \\ -0.24 & 0.06 & 0.14 & -0.08 \\ -0.26 & 0.01 & -0.08 & 0.16 \end{bmatrix}$$

$$(\widetilde{D}^T\widetilde{D} + \alpha I)^{-1} \cdot \widetilde{D}^T = \begin{bmatrix} -0.59 & -0.15 & -0.11 & 0.94 & 0.68 \\ 0.29 & -0.07 & 0.02 & -0.10 & -0.09 \\ 0.04 & -0.13 & 0.31 & -0.06 & -0.13 \\ 0.03 & 0.31 & -0.16 & -0.16 & 0.00 \end{bmatrix}$$

$$\widetilde{w} = (\widetilde{D}^T\widetilde{D} + \alpha I)^{-1} \cdot \widetilde{D}^TY = \begin{bmatrix} -0.267 \\ 0.846 \\ -0.275 \\ 0.622 \end{bmatrix}$$

Bias b is $w_0 = -0.267$

The hyperplane is therefore represented by the fitted model
$Y = -0.267 + 0.846X_1 - 0.275X_2 + 0.622X_3$

Explanation The introduction of an alpha term in ridge regression penalizes the magnitude of the weights in w. To introduce regularization and avoid overfitting, the weights tend to decrease. Analyzing the squared norm, $||w||^2 = 1.42$ for linear regression and decreases to $||w||^2 = 1.25$ in ridge regression. While the squared norm decreases with the introduction of alpha, the bias term, b, increases from -.408 to -.267.

1.c Regularization of b

In this scenario, the bias term should be regularized. Since the data set is very small, not regularizing the bias term can cause the term to adapt too much to the noise since there isn't as much training data to structure it's position. In this situation, the bias term represents a general rating offset of all movies, regardless of any specific characteristic a movie could have, and despite any particular user's singular interests.

2 Decision Trees

2.e Best Splits

As calculated and stated in the code, the best splits are as follows:

GINI (5, 4.0)

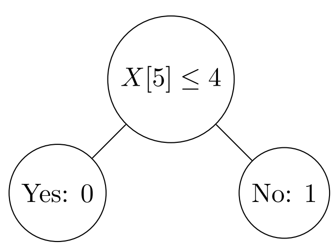
IG (7, 4.0)

CART (2, 0.0)

2.f Tree Diagrams

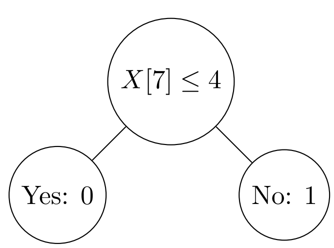
2.f.1 GINI

Values at index 5 for the data sample that are above 4.0 will be classified as 1, others will be classified as 0.



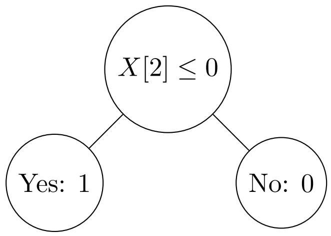
2.f.2 IG

Values at index 7 for the data sample that are above 4.0 will be classified as 1, others will be classified as 0.



2.f.3 CART

Values at index 2 for the data sample that are above 0.0 will be classified as 0, others will be classified as 1.



2.g Error Analysis

The output and error count returned by the program:

GINI [0, 0, 0, 0, 1, 1, 1, 1, 0, 0] Errors: 1

IG [0, 0, 0, 1, 1, 1, 1, 1, 0, 0] Errors: 2

CART [0, 0, 0, 0, 1, 1, 0, 1, 0, 0] Errors: 0

Results The output shows the incorrect classifications in red. It is interesting to note that all misclassifications were of the same class. The program correctly detects errors by comparing each class with the true class.