

Mohammad Javad Ranjbar

810101173

Homework 2

Machine learning, Fall 2022

محمود رشيد ٨١٠١٠١١٧٣

تمرین شماره ٢ یادگیری ماشین

-1

$$R(\theta) = \frac{1}{2N} \|y - X\theta\|_2^2 + \theta^T H \theta + \theta^T \theta + \alpha^T \theta$$

$$\nabla_{\theta} R(\theta) = \nabla_{\theta} \frac{1}{2N} (X\theta - y)^T (X\theta - y) + \nabla_{\theta} \theta^T H \theta + \nabla_{\theta} \theta^T \theta + \nabla_{\theta} \alpha^T \theta$$

$$= \frac{1}{2N} \nabla_{\theta} ((X\theta)^T X\theta - (X\theta)^T y - y^T (X\theta) + y^T y) + 2H\theta + 2\theta + \alpha^T$$

$$= \frac{1}{2N} \nabla_{\theta} (\theta^T (X^T X) \theta - 2(X^T y)^T \theta) + 2H\theta + 2\theta + \alpha^T$$

$$= \frac{1}{2N} (2X^T X\theta - 2X^T y) + 2H\theta + 2\theta = \frac{1}{N} (X^T X\theta - X^T y) + (2H + 2)\theta + \alpha^T$$

$$\nabla_{\theta} R(\theta) = 0 \Rightarrow \theta \left(2H + 2 + \frac{X^T X}{N} \right) = X^T y - \alpha^T$$

$$\Rightarrow \theta = (X^T y - \alpha^T) \left(2H + 2 + \frac{X^T X}{N} \right)^{-1}$$

2- الف)

L1 Regularization:

$$J = \frac{1}{N} \sum_{i=1}^N (y_{\text{predicted}} - y_{\text{original}})^2 + \lambda \sum_{i=1}^n |w_i|$$

معمولاً ویژگی‌ها به هم وابستگی ندارند و می‌توانند مرتبط باشند یا ممکن است ویژگی یک سری از ویژگی‌ها بیشتر یا کمتر از سایر ویژگی‌ها باشد. در L1 Regularization ما به سعی می‌کنیم که جمع وزن‌ها را کوچک‌تری بشود تا از overfit شدن داده‌ها جلوگیری کنیم. در این روش می‌توانیم به صورت کامل ویژگی‌های غیر مرتبط را حذف کنیم.

L2 Regularization:

$$J = \frac{1}{N} \sum_{i=1}^N (y_{\text{predicted}} - y_{\text{original}})^2 + \lambda \sum_{i=1}^n |w_i|^2$$

در روش L2 ما از جمع مربع وزن‌ها استفاده می‌کنیم تا از این overfit شدن جلوگیری کنیم و از ویژگی‌های غیر مرتبط که کمتر

ردش L2	ردش L1
۱- از جمع مربع وزن‌ها استفاده می‌کنیم	۱- جمع قدر مطلق وزن‌ها استفاده می‌کنیم
۲- وزن‌ها در نهایت نزدیک به صفر می‌شوند (Non-sparse)	۲- راه حل وزن‌ها را صفر می‌کنند و متغیرهای ما کاهش می‌یابد (Sparse)
۳- انتخاب ویژگی نداریم	۳- انتخاب ویژگی صورت می‌گیرد
۴- نسبت به داده‌های بی‌ارتباط مقاوم نیست	۴- نسبت به موارد بی‌ارتباط مقاوم است
۵- چندین جواب مختلف ندارد	۵- چندین جواب نمی‌تواند داشته باشد

ب)

$$J = (Y_i - X_i B)^T (Y_i - X_i B) + \lambda B^T B$$

$$Y_i^T Y_i - Y_i^T X_i B - B^T X_i^T Y_i + B^T X_i^T X_i B$$

$$\frac{dJ}{dB} = 0 \Rightarrow 0 - Y_i^T X_i - Y_i^T X_i + 2 X_i^T X_i B + 2 \lambda B \Rightarrow 2 Y_i^T X_i = 2 (X_i^T X_i + \lambda I) B$$

$$\Rightarrow \hat{B} = (X^T X + \lambda I)^{-1} Y^T X_i \Rightarrow (A^T A + \lambda I)^{-1} A^T Y$$

۱۰-

این قانون one against all را نشان می‌دهد که هر بار یک کلاس را در نظریه‌ی بزرگ هر چند آن کلاس نباشد را دلیل دوم قرار می‌دهیم

ب) طبقه‌بندی در هر مرحله باید بزرگترین احتمال را به عنوان کلاس ما انتخاب کند نه عکس

$$\text{class} = \arg \max P(Y = y_k | X)$$

$$y_i = \beta_1 x_i + \beta_0$$

$$J = \sum_{i=1}^n (y_i - \beta_1 x_i - \beta_0)^2 \rightarrow \frac{\partial J}{\partial \beta_1} = 0 \Rightarrow 0 = -2 \sum_{i=1}^n x_i (y_i - \beta_1 x_i - \beta_0)$$

$$\rightarrow \frac{\partial J}{\partial \beta_0} = 0 \Rightarrow 0 = -2 \sum_{i=1}^n (y_i - \beta_1 x_i - \beta_0) \quad (2)$$

$$(1) \Rightarrow \sum_{i=1}^n x_i y_i = \sum_{i=1}^n \beta_1 x_i^2 + \sum_{i=1}^n \beta_0 x_i \quad (3)$$

$$(2) \Rightarrow \sum_{i=1}^n y_i = \sum_{i=1}^n \beta_1 x_i + n \beta_0 \Rightarrow \beta_0 = \frac{1}{n} \sum_{i=1}^n y_i - \frac{1}{n} \beta_1 \sum_{i=1}^n x_i$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x} \quad (4)$$

$$\underline{(3) \& (4)} \rightarrow \beta_1 \sum x_i^2 + (\bar{y} - \beta_1 \bar{x}) \sum x_i = \sum x_i y_i \Rightarrow \beta_1 \sum x_i^2 - \beta_1 \bar{x} \sum x_i + \bar{y} \sum x_i = \sum x_i y_i$$

$$\rightarrow \beta_1 = \frac{\sum x_i y_i - \bar{y} \sum x_i}{\sum x_i^2 - \bar{x} \sum x_i} = \frac{\sum x_i y_i - n \bar{y} \bar{x}}{\sum x_i^2 - n \bar{x}^2}$$

جایگذاری کنیم:

$$\bar{y} = \frac{1}{10} (31 + 58 + 65 + 73 + 37 + 44 + 60 + 91 + 21 + 84) = 56.4$$

$$\bar{x} = \frac{1}{10} (4 + 9 + 10 + 14 + 4 + 7 + 12 + 22 + 1 + 17) = 10$$

$$\sum x_i y_i = 31 \times 4 + 58 \times 9 + 65 \times 10 + 73 \times 14 + 44 \times 37 + 44 \times 7 + 12 \times 60 + 22 \times 91 + 1 \times 21 + 17 \times 84 = 6945$$

$$\sum x_i^2 = 4^2 + 9^2 + 10^2 + 14^2 + 4^2 + 7^2 + 12^2 + 22^2 + 1^2 + 17^2 = 1376$$

$$\beta_1 = \frac{6945 - 10 \times 56.4 \times 10}{1376 - 100 \times 10} = 3.47$$

$$\Rightarrow y = 3.47x + 21.693$$

$$\beta_0 = 56.4 - 3.47 \times 10 = 21.693$$

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

$$\sigma^2 = \text{Var}(\varepsilon_i) = \text{Var}(y_i)$$

متغیر وابسته

$$s^2 = \frac{\sum_{i=1}^n (y_i - y_{\text{mean}})^2}{n-2}$$

فاکتور اری وابسته

$$s^2 = \frac{179.3}{8} = 22.423$$

$$\text{Var}(\beta_1) = \text{Var}\left(\frac{\sum x_i y_i - n \bar{y} \bar{x}}{\sum x_i^2 - n \bar{x}^2}\right) = \left(\frac{1}{\sum x_i^2 - n \bar{x}^2}\right)^2 \text{Var}\left(\sum_{i=1}^n (x_i - \bar{x}) y_i\right)$$

$$= \left(\frac{1}{\sum x_i^2 - n \bar{x}^2}\right)^2 \left(\sum (x_i - \bar{x})\right)^2 \text{Var}(y_i) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{56}{2000}$$

$$\text{Var}(\beta_0) = \text{Var}(\bar{y} - \beta_1 \bar{x}) = \text{Var}(\bar{y}) + \bar{x}^2 \text{Var}(\beta_1) + 2\bar{x} \text{Cov}(\bar{y}, \beta_1)$$

$$\text{Cov}(\bar{y}, \beta_1) = \text{Cov}\left(\frac{1}{n} \sum_{i=1}^n y_i, \beta_1\right) = \frac{1}{n} \frac{1}{\sum x_i^2 - n \bar{x}^2} \text{Cov}\left(\sum_{i=1}^n y_i, \sum_{i=1}^n (x_i - \bar{x}) y_i\right)$$

$$= \frac{1}{n} \frac{1}{\sum x_i^2 - n \bar{x}^2} \sum_{i=1}^n (x_i - \bar{x}) \text{Cov}(y_i, y_i) = 0$$

$$\begin{aligned} \text{Var}(\beta_0) &= \text{Var}(\bar{y}) + \bar{x}^2 \text{Var}(\beta_1) = \frac{\sigma^2}{n \sum_{i=1}^n (x_i - \bar{x})^2} \left(\sum_{i=1}^n (x_i - \bar{x})^2 + n \bar{x}^2\right) = \frac{\sigma^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2} \\ &= 7.67 \end{aligned}$$

ج) برای حل این معادله باید ماتریس کواریانس را محاسبه کنیم

X را به صورت روبروی نویسیم

$$X = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 4 & 9 & 10 & 14 & 7 & 12 & 22 & 1 & 17 \end{bmatrix}$$

$$Cov(B_0, B_1) = \sigma^2 (X^T X)^{-1} = \begin{bmatrix} 10.2 & -0.74 \\ -0.74 & 0.74 \end{bmatrix}$$

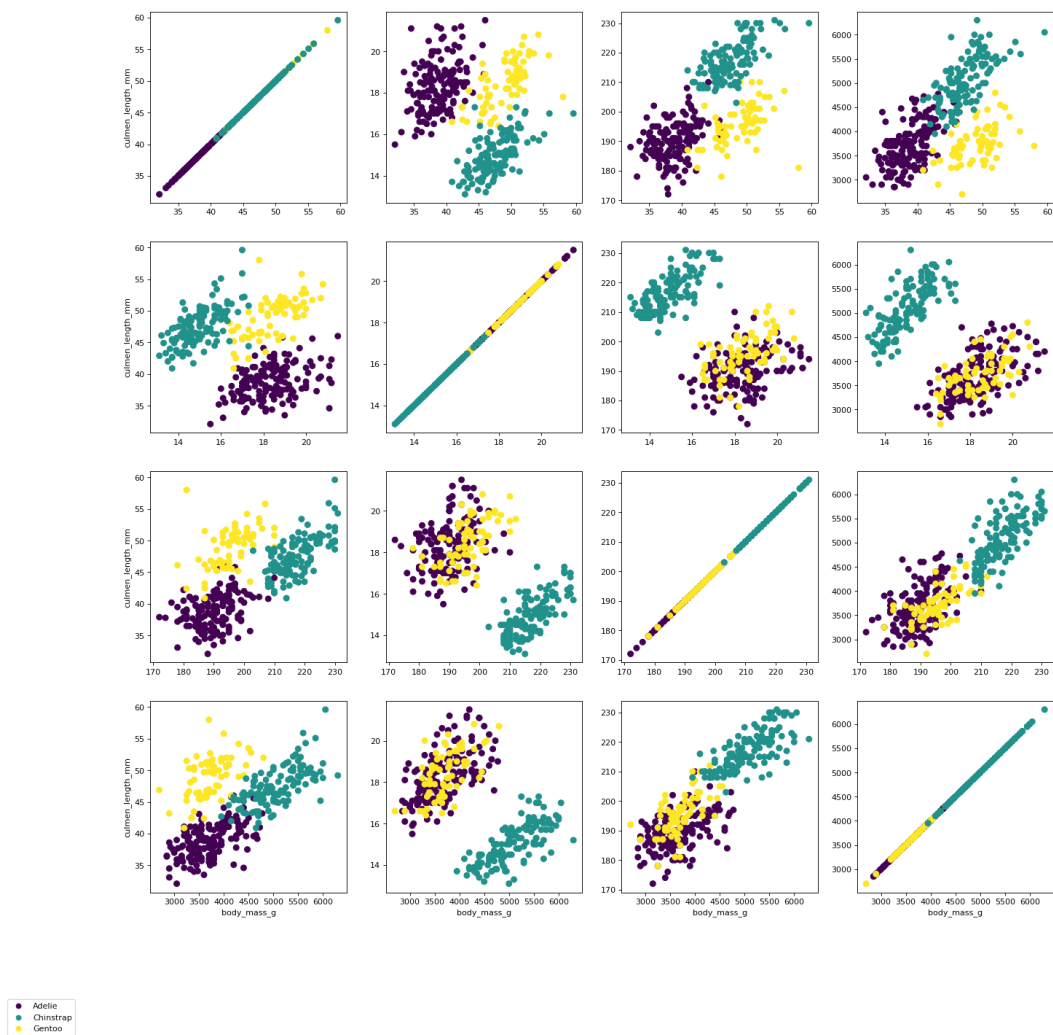
$$Corr(B_0, B_1) = \frac{\sigma_{B_0 B_1}}{\sigma_{B_0} \sigma_{B_1}} = -0.875$$

-5

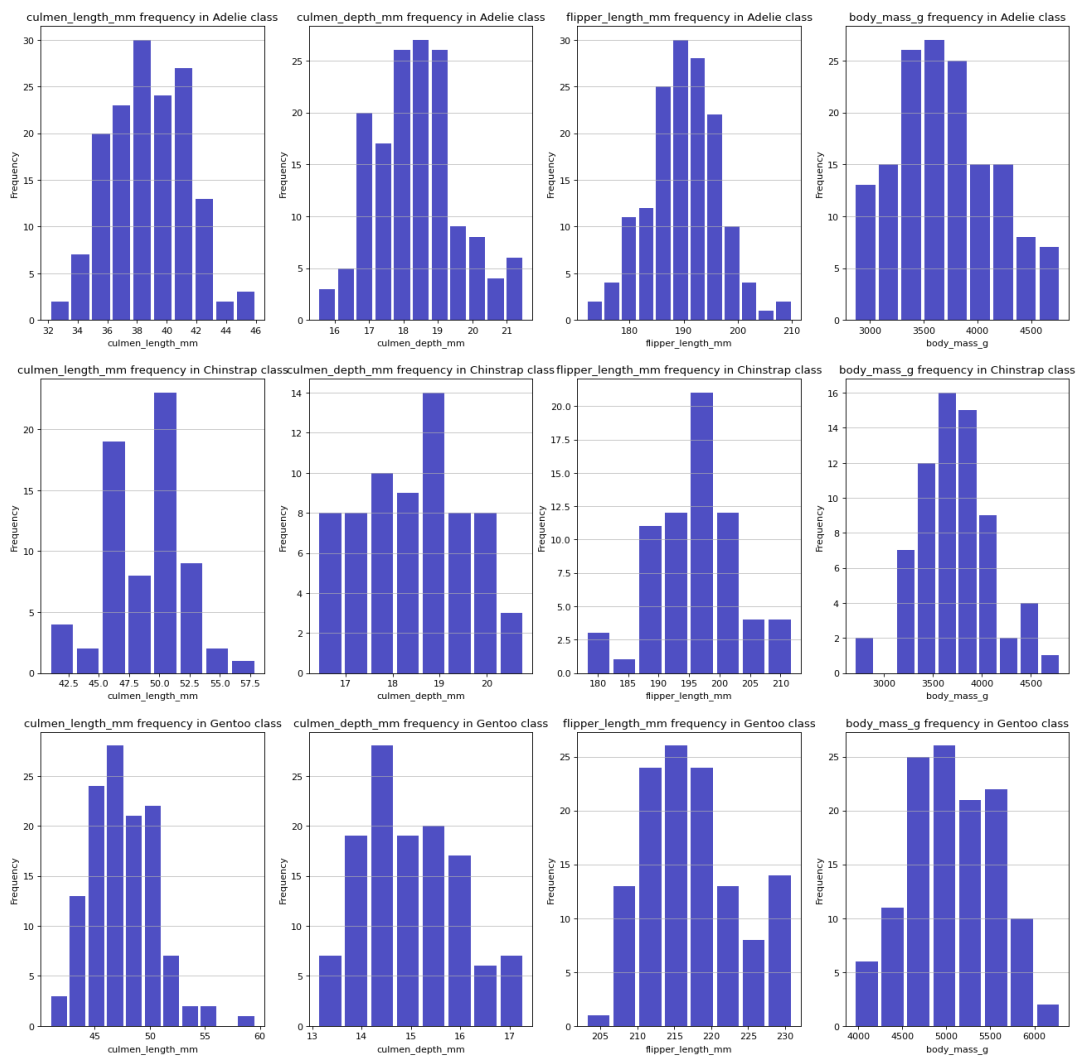
نتایج کاملتر در notebook موجود می باشد.
الف) بر اساس دسته ویژگی های

- 1- culmen_lenght_mm و filipper_lenght_mm
- 2- body_mass_g و culmen_length_mm
- 3- culmen_length_mm و culmen_depth_mm

تا حد خوبی می توان داده ها را به صورت خطی جدا کرد.



هیستوگرام‌ها به شکل زیر می‌باشند.



(ب)

پیاپی سازی بدون استفاده از کتابخانه:

دقت آموزش و تست به صورت زیر می باشد:

Accuracy on train is: 0.9833333333333333

Accuracy on test is: 0.9805825242718447

و به صورت زیر می باشند.

The jaccard score for train data is: [0.9646017699115044,

[0.9285714285714286, 0.9887640449438202

[The precision score for train data is: [0.9646017699115044, 1.0, 1.0

The recall score for train data is: [1.0, 0.9285714285714286,

[0.9887640449438202

The f1-score for train data is: [0.9819819819819819, 0.962962962962963,

[0.9943502824858756

البته می شود برای هر کلاس میانگین هر عدد را اعلام کرد.

کانفیوژن ماتریس ما برای داده آموزش صورت زیر خواهد بود

```
[[109  0  0]
 [ 3 39  0]
 [ 1  0 88]]
```

برای داده تست دقت‌های بالا به صورت زیر خواهد بود.

```
[The jaccard score is: [0.9545454545454546, 0.9230769230769231, 1.0
```

```
[The precision score is: [0.9545454545454546, 1.0, 1.0
```

```
[The recall score is: [1.0, 0.9230769230769231, 1.0
```

```
[The f1-score is: [0.9767441860465117, 0.9600000000000001, 1.0
```

و همچنین کانفیوژن ماتریس زیر را خواهیم داشت:

```
[[15  0  1]
 [ 0 38  1]
 [ 0  0 48]]
```

ج) حال نتایج پیاده سازی با کتابخانه:

```
Accuracy on train data: 0.9875
```

```
Accuracy on test data: 1.0
```

برای داده تست:

```
[The jaccard score for test data is: [0.96078431 0.94444444 0.97297297
```

```
[.The precision score for test data is: [0.96078431 1. 1
```

```
[The recall score for test data is: [1. 0.94444444 0.97297297
```

```
[The f1_score for test data is: [0.98 0.97142857 0.98630137
```

```
array([[49,  0,  0],
       [ 1, 17,  0],
       [ 1,  0, 36]])
```

برای داده آموزش:

```
[The jaccard score is: [0.97115385 0.96078431 0.98863636
```

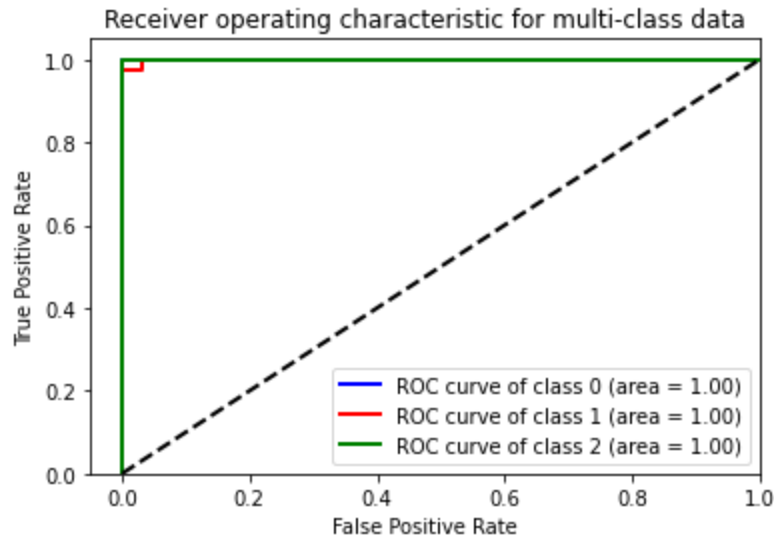
```
[The precision score is: [0.99019608 0.98 0.98863636
```

```
[.The recall score is: [0.98058252 0.98 1
```

```
[The f1_score is: [0.98536585 0.98 0.99428571
```

```
array([[101,  1,  1],
       [ 1, 49,  0],
       [ 0,  0, 87]])
```

برای roc_curve داریم:



و مساحت زیر این نمودار به صورت:

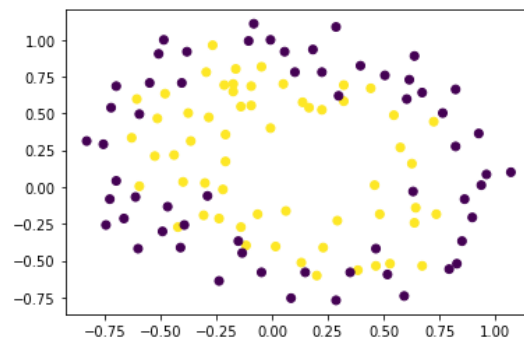
the areas under the curve is {0: 1.0, 1: 0.999231950844854, 2: 1.0}

می‌باشد.

-6

پیاده سازی بدون استفاده از کتابخانه:

داده‌های ما به صورت زیر هستند.



حال به طبقه بند ساخته شد در سوال 5 بخش L2 Regularization را اضافه می‌کنیم که نتایج ما به صورت زیر خواهد شد.

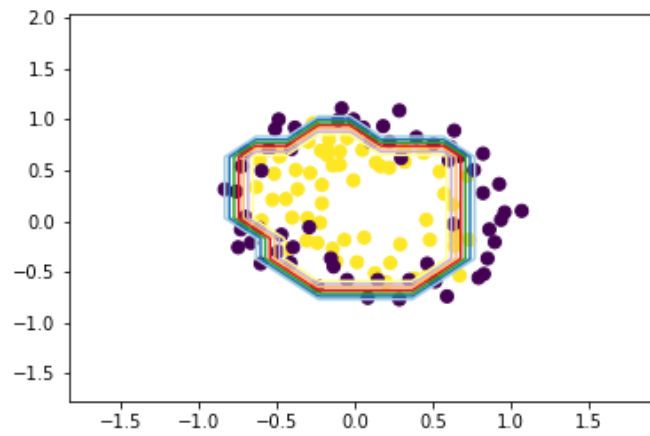
دقت بر روی داده‌های تست و ترین بدون استفاده از کتابخانه به صورت زیر است.

Accuracy for the train data is: 0.8285714285714286

Accuracy for the test data is: 0.8170731707317073

با توجه به اینکه این داده‌ها به صورت خطی جدا پذیر نیستند دقت کمتر این تصمیم گیر نسبت به سوال قبل که خطی جداپذیر بودند قابل توجیه می‌باشد.

و مرز تصمیم گیری ما به صورت زیر خواهد شد:



حال نتایج سازی با استفاده از کتابخانه به صورت زیر خواهد بود:

Accuracy for the train data is: 0.8292682926829268

Accuracy for the test data is: 1.0

ناحیه تصمیم گیری ما برای این طبقه بند به شکل زیر می باشد.

