

تمرین دوم یادگیری عمیق

97521423

محمدعلی فراهت

سوال (1)

- الف) در این سوال محاسبات در قسمت اول نوشته شده اند و در قسمت های بعدی فقط خلاصه ای از آن و مقدار آپدیت شده وزن ها را نشان داده ام.

$$\frac{\partial L}{\partial \hat{y}} = -\frac{1}{N} \sum_{i=1}^N \frac{\hat{y} - y}{\ln(10)(\hat{y}-1) \hat{y}} \quad (الف)$$

$$\frac{\partial \hat{y}}{\partial w_1} = n_1, \quad \frac{\partial \hat{y}}{\partial w_2} = n_2, \quad \frac{\partial \hat{y}}{\partial b} = 1$$

$$\hat{y} = w_1 n_1 + w_2 n_2 + b \quad : \text{batch 1, epoch 1}$$

$$\textcircled{1} \rightarrow = 1 \times 22 + 1 \times 1 + 1 = 24, \quad \textcircled{2} \rightarrow = 1 \times 25 + 1 \times (-1) + 1 = 25$$

$$\text{update: } w'_1 = w_1 - \alpha \frac{\partial L}{\partial w_1} = 1 - 0.05 \left(\frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} \right) = 1$$

$$1 - 0.05 \left(\frac{1}{2} \left(\frac{(24 - (-1))n_1}{\ln_{10}(23)24} + \frac{(25 - (-1))n_2}{\ln_{10}(24)25} \right) \right) = \boxed{0.977}$$

$$w'_2 = w_2 - \alpha \frac{\partial L}{\partial w_2}$$

$$= 1 - 0.05 \left(\frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2} \right) \xrightarrow{\text{محاسبه}} = \boxed{0.999}$$

$$b' = b - \alpha \frac{\partial L}{\partial b}$$

$$= 1 - 0.05 \left(\frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b} \right) \xrightarrow{\text{محاسبه}} = \boxed{0.999}$$

$$\textcircled{3} \rightarrow \hat{y} = 47 \times 0.977 + 1 \times 0.999 + 0.999 = 47.917 \quad : \text{batch 2, epoch 1}$$

$$\textcircled{4} \rightarrow = 52 \times 0.977 + (-1) \times 0.999 + 0.999 = 50.804$$

$$w'_1 = 0.977 - 0.05 \left(\frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} \right) = \boxed{0.955}$$

$$w'_2 = 0.999 - 0.05 \left(\frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2} \right) = \boxed{0.999}$$

$$b' = 0.999 - 0.05 \left(\frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b} \right) = \boxed{0.998}$$

$$⑤ \rightarrow \hat{y} = 0.976 \times 46 + 0.999 \times 1 + 0.998 = 46.893 \quad : \text{batch 3, epoch 1}$$

$$⑥ \rightarrow \hat{y} = 0.976 \times 56 + 0.999 \times 1 + 0.998 = 56.653$$

$$W_1' = 0.976 - 0.05 \left(\frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} \right) = \boxed{0.954} \downarrow$$

$$W_2' = 0.999 - 0.05 \left(\frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2} \right) = \boxed{0.999} \downarrow \rightarrow \text{تغییری جزئی بوده}$$

$$b' = 0.998 - 0.05 \left(\frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b} \right) = \boxed{0.997} \downarrow$$

$$⑦ \rightarrow \hat{y} = 0.954 \times 55 + 0.999 \times (-1) + 0.997 = 54.466 \quad : \text{batch 4, epoch 1}$$

$$⑧ \rightarrow \hat{y} = 0.954 \times 60 + 0.999 \times (-1) + 0.997 = 59.236$$

$$W_1' = 0.954 - 0.05 \left(\frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} \right) = \boxed{0.931} \downarrow$$

$$W_2' = 0.999 - 0.05 \left(\frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2} \right) = \boxed{0.999} \uparrow$$

$$b' = 0.998 - 0.05 \left(\frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b} \right) = \boxed{0.996} \downarrow$$

: batch 1, epoch 2

$$① \rightarrow \hat{y} = 0.931 \times 22 + 0.999 \times 1 + 0.996 = 22.477$$

$$② \rightarrow \hat{y} = 0.931 \times 25 + 0.999 \times (-1) + 0.996 = 23.272$$

$$W_1' = 0.931 - 0.05 \left(\frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} \right) = \boxed{0.906} \downarrow$$

$$W_2' = 0.999 - 0.05 \left(\frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2} \right) = \boxed{0.999} \downarrow$$

$$b' = 0.996 - 0.05 \left(\frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b} \right) = \boxed{0.995} \downarrow$$

$$\textcircled{3} \rightarrow \hat{y} = 0.906 \times 47 + 0.999 \times 1 + 0.995 = 44.576 \quad : \text{batch } \underline{2}, \text{ epoch } \underline{2}$$

$$\textcircled{4} \rightarrow \hat{y} = 0.906 \times 52 + 0.999 \times (-1) + 0.995 = 47.108$$

$$W_1' = 0.906 - 0.05 \left(\frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} \right) = \boxed{0.882} \downarrow$$

$$W_2' = 0.999 - 0.05 \left(\frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2} \right) = \boxed{0.999} \downarrow$$

$$b' = 0.999 - 0.05 \left(\frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b} \right) = \boxed{0.994} \downarrow$$

$$\textcircled{5} \rightarrow \hat{y} = 0.882 \times 46 + 0.999 \times 1 + 0.994 = 42.565 \quad : \text{batch } \underline{3}, \text{ epoch } \underline{2}$$

$$\textcircled{6} \rightarrow \hat{y} = 0.882 \times 56 + 0.999 \times 1 + 0.994 = 51.385$$

$$W_1' = 0.882 - 0.05 \left(\frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} \right) = \boxed{0.858} \downarrow$$

$$W_2' = 0.999 - 0.05 \left(\frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2} \right) = \boxed{0.998} \downarrow$$

$$b' = 0.994 - 0.05 \left(\frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b} \right) = \boxed{0.993} \downarrow$$

$$\textcircled{7} \rightarrow \hat{y} = 0.858 \times 55 + 0.998 \times (-1) + 0.993 = 48.043 \quad : \text{batch } \underline{4}, \text{ epoch } \underline{2}$$

$$\textcircled{8} \rightarrow \hat{y} = 0.858 \times 60 + 0.998 \times (-1) + 0.993 = 51.475$$

$$W_1' = 0.858 - 0.05 \left(\frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} \right) = \boxed{0.832} \downarrow$$

$$W_2' = 0.998 - 0.05 \left(\frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_2} \right) = \boxed{0.998} \uparrow$$

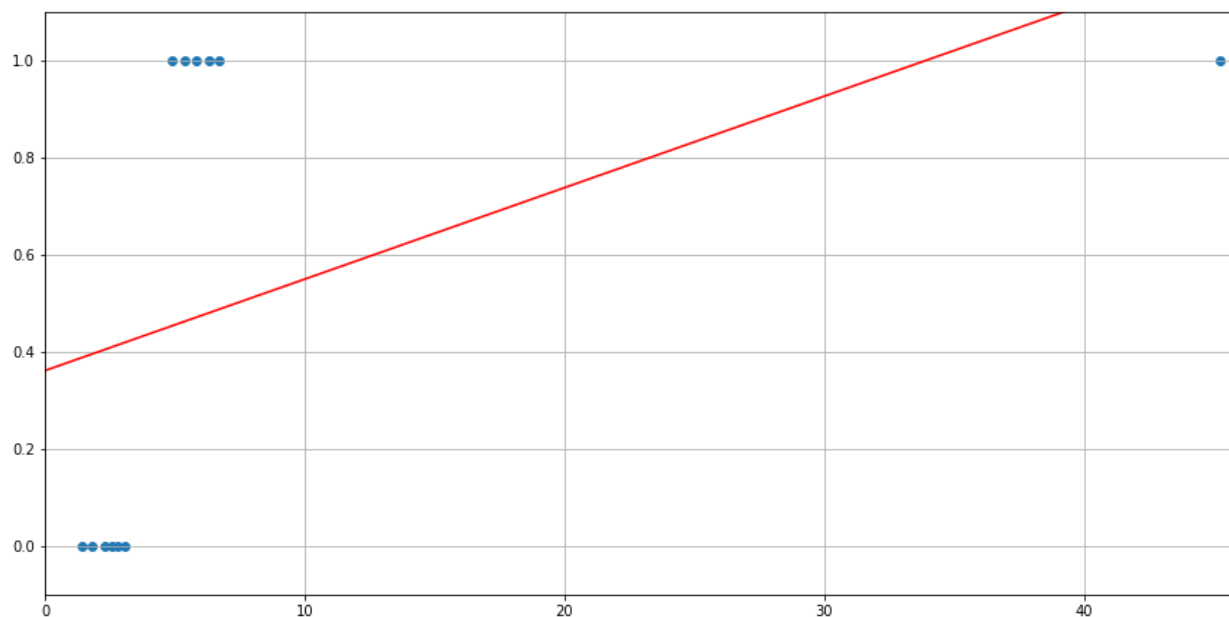
$$b' = 0.994 - 0.05 \left(\frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b} \right) = \boxed{0.992} \downarrow$$

- (ب) در روش گرادیان کاهشی کلاسیک (GD) ما همه دیتا را برایشان forward pass انجام می‌دهیم و سپس برمیگردیم و وزن‌ها را آپدیت می‌کنیم. ولی در روش گرادیان کاهشی تصادفی (SGD) ما هر بار که یک forward pass برای یک ورودی انجام می‌دهیم، بعدش با توجه به loss، وزن‌ها را آپدیت می‌کنیم. در GD اگر دیتا خیلی خیلی بزرگ باشد عملاً این روش اشکالات زیادی دارد، مثلاً اگر وسط محاسبات اشکالی پیش بیاید و محاسبه قطع شود، همه چیز باید دوباره از اول شروع شود. ولی در روش SGD هر جا محاسبات نگه داشته شود، ما وزن‌های آپدیت شده را تا به آنجا داریم. علاوه بر این برای بهبود روش SGD که هزینه محاسباتی بالایی دارد و کند است، می‌توانیم از mini-batch استفاده کنیم. یعنی ترکیبی از هر دو روش که مزیت‌های هر دو را تا حد زیادی پوشش می‌دهد. این یعنی ما دیتا را در batch‌هایی با سایز دلخواه (مثلاً 32، 64، و ...) جمع‌آوری می‌کنیم و در آخر هر batch وزن‌ها را آپدیت می‌کنیم. کاری که در قسمت اول با $\text{batch size} = 2$ انجام دادیم.

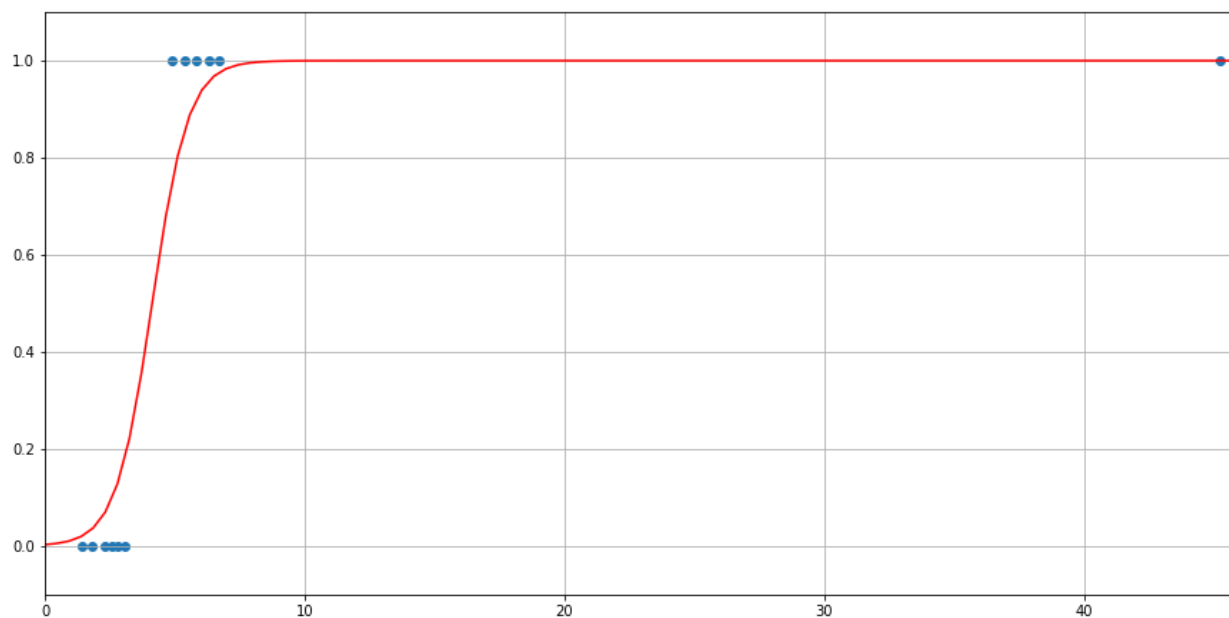
سوال 2) در این سوال ابتدا داده ها را با Linear Regression و سپس با Logistic Regression

آموزش می‌دهیم. سپس نقطه مرزی هر دو را پیدا می‌کنیم. در دو شکل زیر نمودار این دو را می‌بینید:

Linear Regression:



Logestic Regression:



نقطه‌ی مرزی هر کدام را با حل کردن یک معادله ساده به دست می‌آوریم (در کد محاسباتش انجام شده و فقط جواب را اینجا نشان میدهم) :

Linear Regression:

Boundary point is : (7.358333333333335, 0.5)

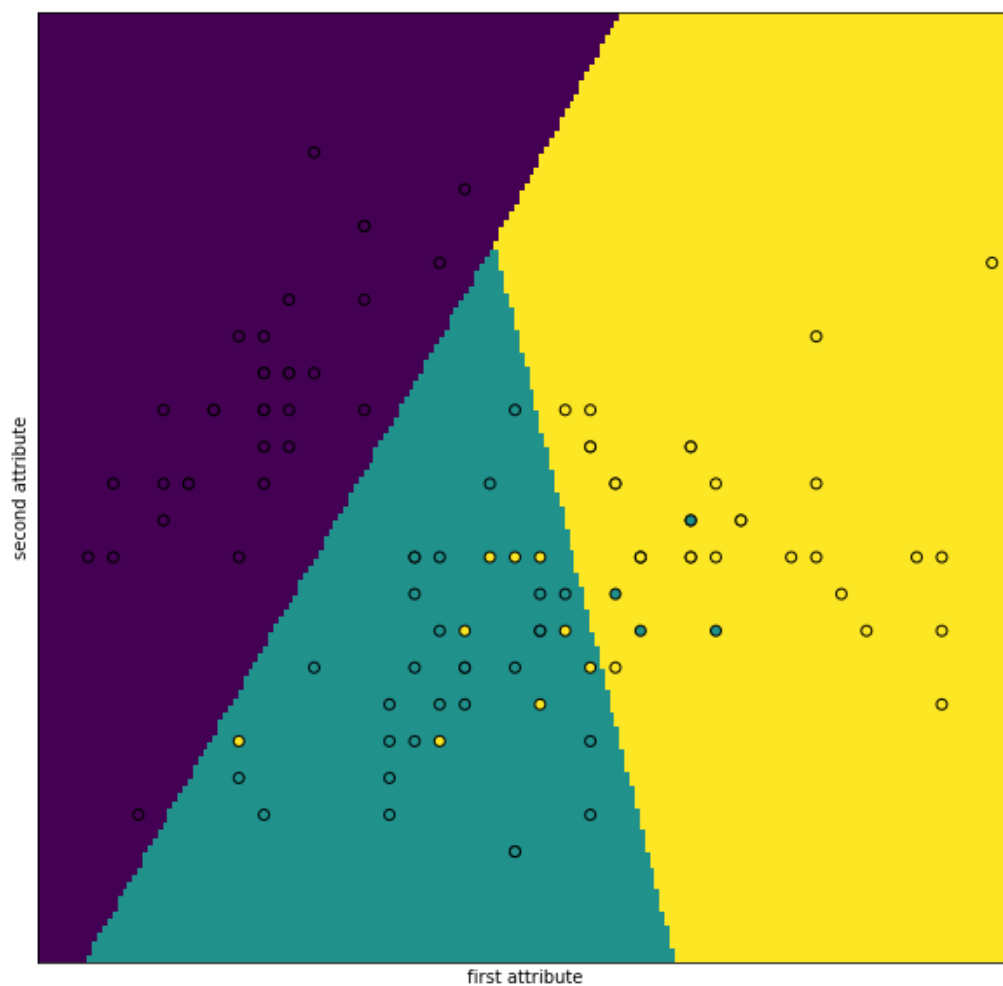
Logestic Regression:

Boundary point is : (4.128272972239653, 0.5)

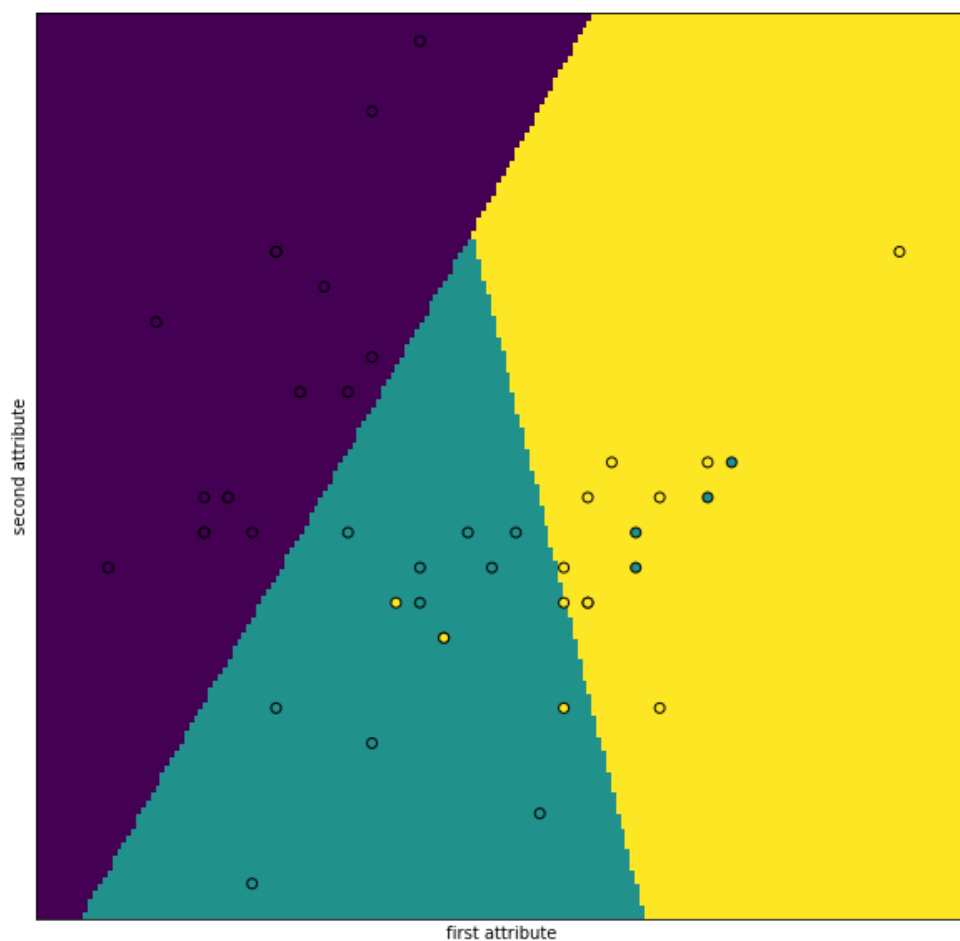
- در Linear regression میبینیم که وقتی نقاط سمت چپ در کلاس $y=1$ ها ، فشرده میشوند و خیلی نزدیک کلاس $y=0$ ها میشوند ، دقت پایین می آید و خیلی از نقاط را اشتباه predict خواهد کرد. اما در Logistic regression این مشکل کمتر است و نه تنها نقاط فشرده سمت چپ را درست طبقه بندی می کند ، بلکه نقطه سمت راست که خیلی فاصله دارد با بقیه و در linear regression اشتباه طبقه بندی میشد ، را نیز درست جدا میکند و این نشان میدهد که Logistic regression بیشتر اوقات بهتر است.

سوال 3) در این سوال به چند بخش جواب میدهم:

- الف) مجموعه داده iris، یک دیتاست جمع آوری شده درباره چند نوع گل است. تعداد attribute هایی که برای هر گل وجود دارد 4 تا است. در کل اطلاعات 150 مورد گل وجود دارد که آن ها به 3 دسته تقسیم میشوند. در این داده ها تست و آموزش جدا نشده و خودمان باید آن را جدا کنیم.
- ب) این نمای دو بعدی از داده های training است. هر رنگ نشان دهنده ی یک کلاس است.



- ج) نمای دو بعدی داده های آزمون (30٪ کل دیتا)



- د) دقت فاز آموزش (در آخرین اجرای برنامه) : 82.5٪

دقت فاز آزمون (در آخرین اجرای برنامه) : 82.2٪

در مقایسه این اطلاعات به این نتیجه رسیدم که دقت آموزش و تست خیلی نزدیک هستند اکثر اوقات دقت تست ها کمتر از آموزش است و این منطقی هم هست.

- ه) confusion matrix را برای هر دو دیتای تست و آموزش تشکیل دادم:

```
training confusion matrix:
[[33  0  0]
 [ 0 28  8]
 [ 0 10 26]]
test confusion matrix:
[[17  0  0]
 [ 0 10  4]
 [ 0  4 10]]
```

در این ماتریس ها ما داریم اختلاف های مقدار های پیشبینی شده را با مقدار های واقعی در یک ماتریس نشان می‌دهیم ، هرچه قطر اصلی اعداد بزرگتری داشته باشد به نسبت بقیه ، یعنی دقت بالاتر است. مثلا در شکل زیر عدد 10 یعنی مدل ما 10 بار اشتباه کرده و به جای گزارش عدد 3 به انواع جواب ، عدد 2 گزارش شده .

```
training confusion matrix:
[[33  0  0]
 [ 0 28  8]
 [ 0 10 26]]
```

با توجه به نگاه کردن به اعداد ماتریس میتوان اینطور تحلیل کرد که label های 2 و 3 خیلی اوقات (هم در داده تست و هم در آموزش) با هم اشتباه گرفته میشوند . و این ها میتوانند همان نقاطی باشند که در کنار مرز سبز و زرد در تصاویر قسمت الف و ب هستند و همانطور که در آن تصاویر میبینیم بنفش ها زیاد از مرز بیرون نرفته اند و برای همین اینجا هم خانه های آن ها صفر است.

لینک منابع استفاده شده (همه از سایت scikit-learn.org هستند) :

استفاده از Logistic regression

جداسازی دیتای تست و آموزش