

۱- الف)

محاسبات را به ازای یک batch به صورت دستی انجام داده و بقیه محاسبات با استفاده از کد پایتون که در نوت بوک موجود است انجام میشود.

Epoch : 1

Batch = 2 -> data set: [(22, 1, 0), (25, 0, 0)]

ابتدا تابع سیکموید رو برای هر دو دیتاپوینت محاسبه میکنیم:

$$z_1 = w_1 * x_1 + w_2 * x_2 + b = 22 + 1 + 1 = 24$$

$$\hat{y}_1 = \frac{1}{(1 + e^{-z})} = 0.9999999999622486$$

$$z_1 = 25 + 0 + 1 = 26$$

$$\hat{y}_2 = 0.999999999994891$$

سپس loss function را محاسبه میکنیم:

$$L = -\frac{1}{2} * (0 * \log(\hat{y}_1) + 1 * \log(1 - \hat{y}_1) + 0 * \log(\hat{y}_2) + 1 * \log(1 - \hat{y}_2))$$

$$= 25.000006172241022$$

حال باید با استفاده از sgd وزن ها رو آپدیت کنیم. برای این کار لازم است گرادیان تابع ضرر رو نسبت به وزن ها بدست آوریم.

$$\frac{\delta L}{\delta \hat{y}} = -\frac{1}{N} * \sum \frac{y}{\hat{y}} + \frac{1 - y}{1 - \hat{y}}$$

$$\frac{\delta \hat{y}}{\delta z} = -\frac{1}{(1 + e^{-z})^2} * -e^{-z} = \frac{e^{-z} + 1 - 1}{1 + e^{-z}} * \frac{1}{1 + e^{-z}} = \hat{y} * (1 - \hat{y})$$

$$\begin{aligned}\frac{\delta L}{\delta w_1} &= \frac{\delta L}{\delta \hat{y}} * \frac{\delta \hat{y}}{\delta z} * \frac{\delta z}{\delta w_1} = -\frac{1}{N} * \sum \left(\frac{y}{\hat{y}} + \frac{y-1}{1-\hat{y}} \right) * \hat{y} * (1-\hat{y}) * x_1 \\ &= -\frac{1}{2} * ((y_1 - \hat{y}_1) * 22 + (y_2 - \hat{y}_2) * 25) \\ &= 23.499999999520874\end{aligned}$$

$$\frac{\delta L}{\delta w_2} = 0.4999999999811243$$

$$\frac{\delta L}{\delta b} = 0.9999999999785698$$

حال با استفاده از گرادیان های محاسبه شده، w ها را آپدیت میکنیم:

$$w_1 = w_1 - 0.05 * \frac{\delta L}{\delta w_1} = -0.17499999997604365$$

$$w_2 = w_2 - 0.05 * \frac{\delta L}{\delta w_2} = 0.9750000000009438$$

$$b = b - 0.05 * \frac{\delta L}{\delta b} = 0.9500000000010715$$

و به همین ترتیب برای بقیه $batch$ و $epoch$ ها w ها را محاسبه میکنیم:

$$w_1: 1, w_2: 1, b: 1$$

$$e: 1, batch: 1 \rightarrow w_1: -0.175, w_2: 0.975, b: 0.95$$

$$e: 1, batch: 2 \rightarrow w_1: 0.997, w_2: 0.999, b: 0.975$$

$$e: 1, batch: 3 \rightarrow w_1: 0.997, w_2: 0.999, b: 0.975$$

$$e: 1, batch: 4 \rightarrow w_1: -0.377, w_2: 0.999, b: 0.95$$

$$e: 2, batch: 1 \rightarrow w_1: -0.378, w_2: 0.999, b: 0.95$$

$$e: 2, batch: 2 \rightarrow w_1: 0.796, w_2: 1.024, b: 0.974$$

$$e: 2, batch: 3 \rightarrow w_1: 0.796, w_2: 1.024, b: 0.974$$

$$e: 2, batch: 4 \rightarrow w_1: -0.578, w_2: 1.024, b: 0.949$$

(۱- ب)

۱. در گرادین کاهشی عادی یا دسته ای (batch gradient descent)، با کل سمپل ها train میکنیم ولی در روش تصادفی از یم سمپل استفاده میکنیم.
۲. گرادین کاهشی بسیار آرام و هزینه بر است. برای این مشکل میتوان از روش های بهینه سازی ای مانند adam استفاده کرد.
۳. در حالت عادی هیچ شافل رندومی روی نقاط لازم نیست ولی در روش تصادفی نیاز داریم که در هر epoch، سمپل به صورت رندوم باشد پس لازم است که روی آن شافل بزنیم.
۴. در روش تصادفی میتوان به راحتی از مینیموم های محلی خارج شد ولی روش عادی خیلی وقت ها در local minimum ها گیر میکند.
۵. در روش عادی مدل به سختی همگرا میشود ولی در روش تصادفی سرعت این اتفاق بیشتر است که برای آن میتوان که learning rate داینامیک داشته باشیم که به همگرایی سریع تر آن کمک کند.

منبع:

<https://www.geeksforgeeks.org/difference-between-batch-gradient-descent-and-stochastic-gradient-descent/>

۲-

دو مدل Logistic Regression و Linear Regression با استفاده از کتابخانه sklearn درون جویتر نوت بوک پیاده سازی شده اند.

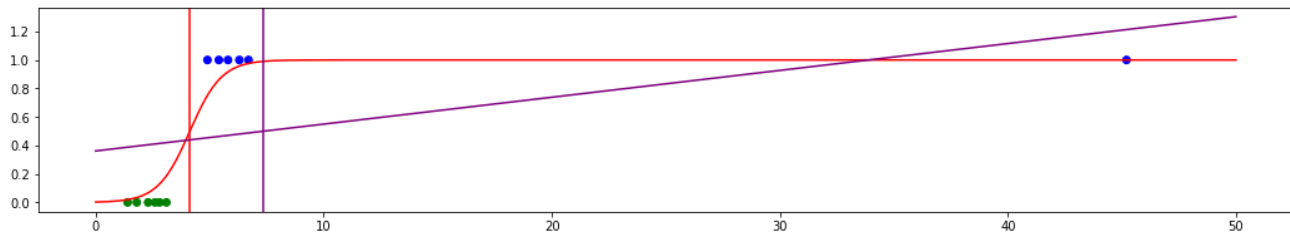
در ابتدا دیتا ست را درست کرده و هر دو مدل را درون سل های جدا فیت کرده و w و b هر کدام را بدست میاوریم. مرز تصمیم گیری برای مدل Logistic Regression زمانی است که تابع سیگموئید که تابع فعال ساز این مدل است برابر ۰.۵ باشد:

$$\frac{1}{1 + e^{-z}} = \frac{1}{2} \rightarrow e^{-z} = 1 \rightarrow z = 0 \rightarrow wx + b = 0 \rightarrow x = \frac{-b}{w}$$

برای Linear Regression به شکل به صورت زیر مرز تصمیم را پیدا کردیم:

$$wx + b = \frac{1}{2} \rightarrow x = \frac{0.5 - b}{w}$$

در آخر نمودار مربوط به دو مدل را در یک پلات کشیده تا بهتر مقایسه شوند.



همانطور که میبینیم، Linear Regression عملکرد خوبی در این سوال نداشته است.

یکی از تفاوت های این دو مدل این است که Linear Regression تخمینش یک عدد است ولی در logistic فقط ۰ یا ۱ پیشبینی میشود. به همین دلیل در مسائلی که جواب پیوسته دارند مانند تخمین قیمت و ... بیشتر از Linear Regression و در مسائل گسسته مانند این مسئله از logistic استفاده میشود.

از آنجایی که Linear Regression براساس فاصله تخمین میزند در مسائلی از این دست نمیتواند خوب عمل کند ولی در logistic که بر اساس maximum likelihood است عملکردش بهتر استم.

منبع:

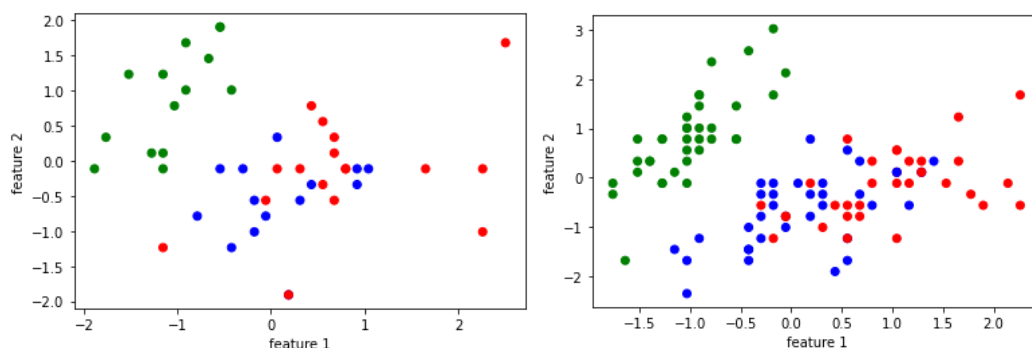
<https://www.geeksforgeeks.org/ml-linear-regression-vs-logistic-regression/>

۳-

دیتاست iris یک مجموعه داده چند متغیری از گل های زنبق است که توسط Ronald Fisher در مقاله ۱۹۳۶ منتشر شد که از آن برای اندازه گیری های مختلف در مسائل classification استفاده میشود.

این دیتاست شامل ۱۵۰ گل و از هر کدام ۴ فیچر است و در سه دست لیبل خورده اند.

نمای آزمون داده های train و test که در نوت بوک نیز موجود است:



دقت در فاز آموزش تقریباً برابر با ۸۰ درصد بود که در فاز تست نیز تقریباً برابر با همین مقدار بود که نشان دهنده این است که مدل ما **over-fit** نشده است.

در نهایت نیز با مشاهده **confusion matrix** میتوان دید که داده های با لیبل ۰ با دقت ۱۰۰ درصد پیدا شده اند ولی مدل کمی در تشخیص لیبل های ۱ و ۲ از هم مشکل دارد و خطا دارد.

