

دسته بندی پرسپترون Perceptron

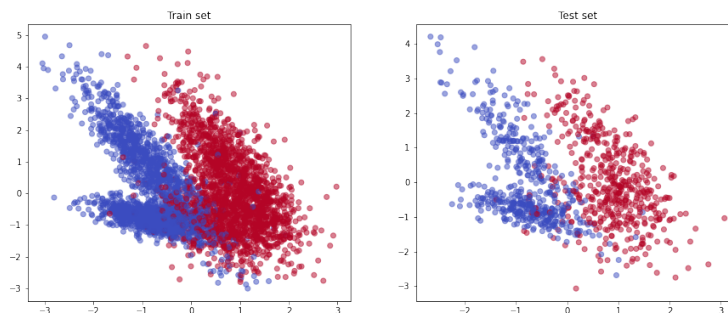
سیدعلیرضا مولوی

فهرست مطالب

۲	۱ بررسی داده ها
۲	۲ الگوریتم دسته بندی
۳	۳ مدل آموزش داده شده

۱ بررسی داده ها

طبق شکل ۱ به نظر می آید میتوان داده ها رو با یک خط از یکدیگر جدا کرد؛ ولی با کمی دقت متوجه می شویم که بخشی از داده های کلاس با یکدیگر تداخل دارند که احتمالا بعدا مشکل ساز میشوند.
نکته: ما فقط باید داده های train بررسی کنیم و مدل را بر اساس آن بسازیم و از داده های test فقط برای ارزیابی نهایی مدل استفاده کنیم.



شکل ۱: داده های train در سمت چپ قرار و داده های test در سمت راست قرار دارند.

۲ الگوریتم دسته بندی

با توجه به اینکه داده ها نسبتا به طور خطی قابل جدا سازی اند از الگوریتم پرسپترون^۱ استفاده میکنیم. پرسپترون یک الگوریتم دسته بندی خطی است؛ بنابراین قابلیت تقسیم داده های غیر خطی را ندارد. فرضیه را به صورت معادله ۱ می سازیم. که X ورودی و Y برچسب ها و \hat{Y} خروجی مدل به ازای ورودی است و W, b متغیرهای مدل اند. هدف این است که مقادیر \hat{Y} و Y تا حد مشابه باشند.

متغیرها	بعد
X	$(N \times D)$
Y, \hat{Y}	$(N \times 1)$
W	$(D \times 1)$
b	$(1,)$

$$\hat{Y} = \text{sign}(X.W + b) \quad (۱)$$

جدول ۱: بعد متغیرها

مقادیر اولیه W, b در ابتدا به صورت تصادفی انتخاب میشوند. در کد نوشته شده $W \sim N(\mu = 0, \sigma = 1)$ و $b = 0$ در نظر گرفته شده اند.

در کد نوشته شده برای آموزش مدل از روش mini batch استفاده شده است. $x^{(i)}, y^{(i)}$ به طور تصادفی از دیتاست انتخاب شده اند (l تعداد mini batch است).
 طریقه آموزش متغیرها در معادلات شکل ۲ نوشته شده است. معادلات ذکر شده رو به طور متوالی انجام میدهم و این باعث میشود که مدل به تدریج خطی که کلاس ها رو از یکدیگر جدا میکند را بیابد.

^۱Perceptron

$$W = W - learning_rate \times \Delta_w \quad (2)$$

$$b = b - learning_rate \times \Delta_b \quad (3)$$

$$\Delta_w = \frac{1}{l} \sum_{i=1}^l (\hat{y}^{(i)} - y^{(i)}) x^{(i)} \quad (4)$$

$$\Delta_b = \frac{1}{l} \sum_{i=1}^l (\hat{y}^{(i)} - y^{(i)}) \quad (5)$$

شکل ۲: معادلات یادگیری

۳ مدل آموزش داده شده

هایپرپارامتر^۲ ها را $learning_rate = 0.0001$, $batch_size = 256$, $epochs = 100000$ قرار دادیم و مدل را آموزش دادیم.

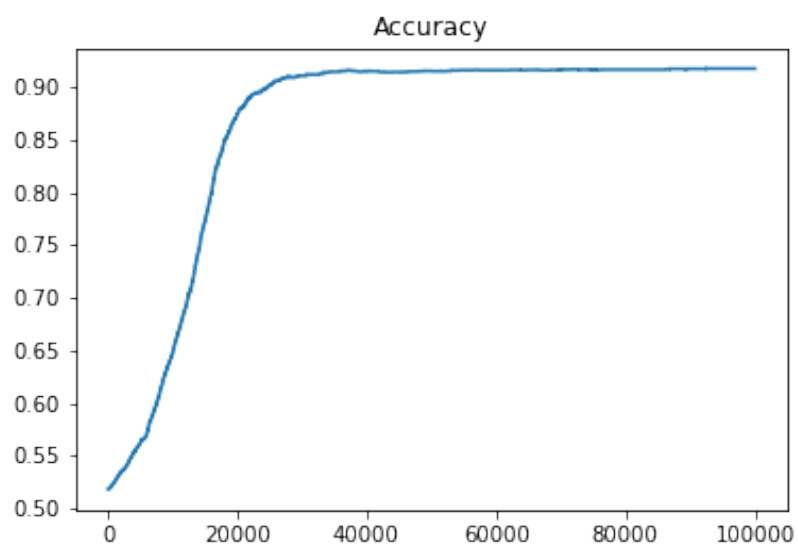
دقت بر روی داده های آموزش = 0.927 دقت بر روی داده های تست = 0.91825

دقت مدل در طول آموزش در شکل ۳ مشاهده میکنیم.

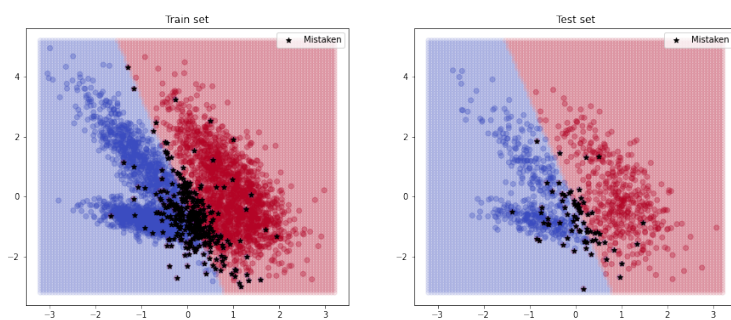
نکته مهم: نمودار میتواند بسته به مقدار دهی اولیه متغیرهای W و b و برای مقادیر متفاوت هایپرپارامترها متفاوت باشد؛

اما اگر نرخ یادگیری کوچک باشد و الگوریتم برای دورهای زیاد تکرار شود تقریباً به نتیجه یکسانی همگرا میشود.

در شکل ۴ عملکرد مدل بر روی داده مشاهده میکنید. الگوریتم توانسته خطی رو پیدا کند که داده ها رو به دو بخش تقسیم کند. ولی یک سری داده (داده های مشکی رنگ) در شکل ۴ به اشتباه دسته بندی شده اند. این داده ها عموماً بر نزدیکی خط جدا کننده قرار دارند. دلیل وجود داده ها میتواند نویز یا اشتباه سیستم جمع کننده داده باشد.



شکل ۳: دقت مدل در طول آموزش



شکل ۴: مدل آموزش داده شده. داده های سیاه رنگ داده هایی اند که به اشتباه دسته بندی شده اند.