

شبکه‌های عصبی و یادگیری عمیق
دکتر صفابخش



دانشگاه صنعتی امیرکبیر

(پلی تکنیک تهران)

دانشکده مهندسی کامپیوتر

رضا آدینه پور ۴۰۲۱۳۱۰۵۵

تمرین اول

شبکه Perceptron و Adalin

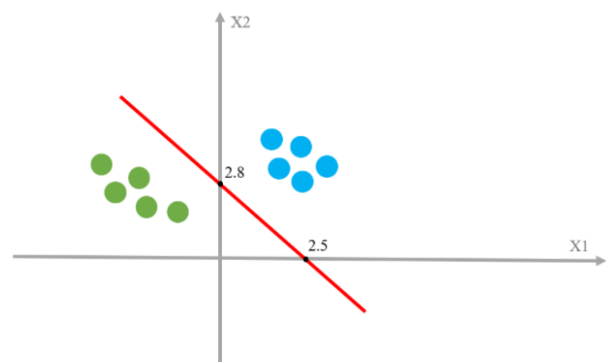
۲ فروردین ۱۴۰۳

سوال اول - نظری

همانگونه که در کلاس درس آشنا شده‌اید، واحد پردازشی پرسپترون و آدالین امکان دریافت ورودی، توان‌های متعدد آن و حاصل ضرب ورودی‌ها را داشته و می‌تواند مسئله دسته‌بندی خطی را حل نمایند. در این سوال، قصد بدست آوردن وزن‌های یک نرون پردازشی پرسپترون را به صورت نظری و با محاسبات دستی داریم.



شکل ۱-ب



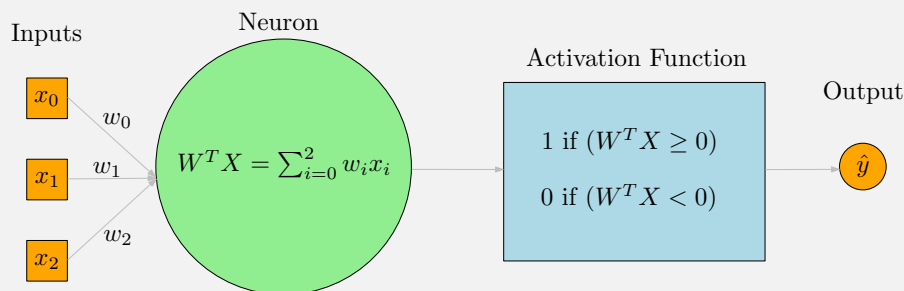
شکل ۱-ا

شکل ۱: مسئله مورد بحث

۱. شکل ۱-ا را برای دسته بندی مسئله دودویی در نظر بگیرید. معماری نرون مورد نظر را توضیح داده و وزن‌های آن را بدست آورید.

پاسخ

به دلیل آنکه داده های این قسمت جداپذیر خطی هستند، می‌توان برای طبقه بندی آنها، از شبکه پرسپترون تک لایه استفاده کرد. معماری این شبکه به صورت «شکل ۲» است.



شکل ۲: معماری شبکه پرسپترون تک لایه

پاسخ

در این شبکه، ورودی/خروجی‌ها با مربع‌های نارنجی، نورون‌ها با دایره سبز و تابع فعال ساز با مربع آبی نشان داده شده است. تعداد دیتا ورودی شبکه ۲ است. x_1 و x_2 . بایاس این شبکه با x_0 نشان داده شده است. وزن‌های شبکه نیز با w نشان داده شده است. بنابراین بردار ورودی و وزن‌های شبکه به صورت زیر است:

$$X = \begin{bmatrix} x_0 = 1 \\ x_1 \\ x_2 \end{bmatrix} \quad W = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix}$$

طبق تئوری شبکه‌های عصبی می‌دانیم خروجی نرون به صورت زیر محاسبه می‌شود: (در اینجا برای انجام محاسبات ساده، تابع فعال‌ساز در نظر گرفته نشده است)

$$\hat{y} = W^T X = \sum_{i=0}^2 w_i x_i = w_0 x_0 + w_1 x_1 + w_2 x_2 \xrightarrow{x_0=1} w_0 + w_1 x_1 + w_2 x_2$$

طبق شکل a-1 دو نقطه از خط جداکننده دو کلاس را داریم. بنابراین می‌توان معادله خط را به صورت زیر نوشت. می‌دانیم معادله خط به صورت زیر تعریف می‌شود:

$$y - y_0 = m(x - x_0)$$

که در آن m شیب خط است و به صورت زیر $\frac{\Delta y}{\Delta x}$ تعریف می‌شود. با جاگذاری یک از نقاط در معادله خط، می‌توان معادله خط را بدست آورد.

$$P_1 = \begin{bmatrix} 2.5 \\ 0 \end{bmatrix} \quad P_2 = \begin{bmatrix} 0 \\ 2.8 \end{bmatrix}$$

$$m = \frac{2.8 - 0}{0 - 2.5} = -1.12 \rightarrow y - 0 = -1.12(x - 2.5) \rightarrow \boxed{y = -1.12x + 2.8}$$

حالا اگر معادله خروجی نورون را به صورت زیر مرتب کنیم، می‌توان از مقایسه با معادله خط بدست آمده وزن‌های شبکه را تعیین کرد.

$$x_1 = \frac{-w_2}{w_1} x_2 - \frac{w_0}{w_1}, \quad x_2 = \frac{-w_1}{w_2} x_1 - \frac{w_0}{w_2}$$

در اینجا به دلیل آنکه دو معادله و ۳ مجهول (w_0, w_1, w_2) داریم، نیاز است که یکی از وزن‌ها را فرض کرده و دو وزن دیگر را بدست آورد.

$$\begin{aligned} \frac{-w_2}{w_1} &= -1.12 \rightarrow w_2 = 1.12w_1 \\ \frac{-w_0}{w_1} &= 2.8 \rightarrow w_0 = -2.8w_1 \\ \rightarrow \begin{cases} w_2 - 1.12w_1 = 0 \\ -2.8w_1 - w_0 = 0 \end{cases} & \text{assume } w_0 = 2.8 \rightarrow \boxed{w_1 = -1, w_2 = -1.12} \end{aligned}$$

ذکر این نکته الزامیست که این جواب، یکتا نمی‌باشد و برحسب اینکه مقدار w_0 را چه انتخاب کنیم، مقدار ۲ وزن دیگر متفاوت می‌شود.

۲. حال شکل b-1 را در نظر بگیرید. چرا مسئله جداپذیر خطی نیست؟ چگونه می‌توان آن را در قالب حل چند مسئله خطی حل نمود؟ معماری پیشنهادی خودتان را رسم و وزن‌های موجود در آن را با انجام محاسبات بدست آورید. معماری شما می‌تواند حاصل از کنار هم چیدن و پشت هم چیدن یک یا چند نورون پرسپترون باشد.

پاسخ

به مسائلی جداپذیر خطی گفته می‌شود که بتوان داده‌ها (کلاس‌ها) را با استفاده از فقط یک خط از هم جدا کرد. در این مثال، دو کلاس آبی و سبز را نمی‌توان فقط با یک خط از هم جدا کرد. بنابراین این مسئله جداپذیر خطی نمی‌باشد. برای حل این مسئله، چندین روش وجود دارد که در ادامه آنها را توضیح خواهیم داد.

(آ) افزایش ابعاد ویژگی‌های مسئله:

یعنی در این مثال که مسئله مورد بحث ما دو بعدی است، یک بعد به آن اضافه کنیم و آن را به عنوان یک مسئله ۳ بعدی حل کنیم و تلاش کنیم که یک صفحه برای جدا کردن دو کلاس پیدا کنیم.

(ب) دادن ورودی‌هایی با توان بالا به عنوان ورودی شبکه:

در صورتی می‌توان از این روش استفاده کرد که مقدار دقیق تمام نمونه‌ها را داشته باشیم. در این صورت می‌توان ورودی‌ها را به توان‌های بالا (۲ و ۳ و...) رساند و امیدوار باشیم فضای قرارگیری ویژگی‌های جدید در صفحه به صورت جداپذیر خطی باشد. «این کار در سوال دوم همین سری تمرین انجام شده است و با به توان ۲ رساندن ویژگی‌های ورودی شبکه، مسئله‌ای که جداپذیر خطی نبود به جداپذیر خطی تبدیل می‌شود.» در این مسئله به دلیل نداشتن موقعیت دقیق نمونه‌های هرکلاس نمی‌توان از این روش استفاده کرد.

(ج) اضافه کردن لایه مخفی:

در اضافه کردن تعداد لایه‌های مخفی شبکه، آزاد هستیم اما باید به این نکته توجه داشت که اگر بیشتر از یک لایه مخفی به شبکه اضافه کنیم، حل مسئله از نظر خطی بودن خارج شده و نواحی تصمیم‌گیری شامل خط راست نمی‌شود و نواحی پیچیده‌تری مانند منحنی‌ها را در بر می‌گیرد. اما اگر فقط یک لایه مخفی ۳ نورونی به شبکه اضافه کنیم، می‌توان مسئله‌ای که ذاتاً جداپذیر خطی نیست را به وسیله ۳ خط جدا کننده که تعداد این خط‌ها برابر است با تعداد نورون‌های لایه مخفی، حل نمود. ساختار این مدل در «شکل ۳» آورده شده است.

اما مشکلی که وجود دارد آن است که در صورت سوال از ما خواسته شده وزن‌های شبکه را بدست آوریم، اگر از این ساختار برای حل استفاده کنیم، به دلیل نداشتن مقدار ورودی‌ها نمی‌توان ۳ وزن لایه متصل به خروجی را بدست آورد.

(د) شکستن مسئله به ۳ زیرمسئله خطی:

برای حل این سوال از این روش استفاده شده است. در قسمت قبل دیدیم که یک نورون پرسپترون می‌تواند یک خط جدا کننده در صفحه رسم کند. در این مثال، برای جدا کردن این دو کلاس به ۳ خط نیاز داریم. پس داده‌های ورودی مسئله را به ۳ بخش جداپذیر خطی تقسیم می‌کنیم. «شکل ۴» اکنون می‌توان همانند قسمت قبل، وزن‌ها را برای هر سه زیرمسئله بدست آورد. در این قسمت ساختار و معماری شبکه همان ساختار قسمت a-1 است. «شکل ۲»

پاسخ



شکل ۳: معماری شبکه ۳ لایه پیشنهادی



شکل ۴: داده‌های شکسته شده به سه بخش

زیر مسئله شماره ۱ در قسمت قبل حل شده است و وزن‌های آن به صورت زیر بدست آورده شد:

$$w_0 = 2.8, \quad w_1 = -1, \quad w_2 = -1.12$$

برای دو زیر مسئله دیگر هم، همانند قسمت قبل وزن‌ها را بدست می‌آوریم.

$$P_1 = \begin{bmatrix} 2.5 \\ 0 \end{bmatrix} \quad P_2 = \begin{bmatrix} -1.5 \\ 1.7 \end{bmatrix} \rightarrow m = \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1} = \frac{-1.5 - 2.5}{-1.7 - 0} = \frac{-4}{-1.7} = \boxed{2.35}$$

$$y - 0 = 2.35(x - 2.5) \rightarrow \boxed{y = 2.35x - 5.88}$$

$$\rightarrow x_1 = \frac{-w_2}{w_1}x_2 - \frac{w_0}{w_1} \rightarrow \begin{cases} \frac{-w_2}{w_1} = 2.35 \rightarrow w_2 = -2.35w_1 \\ \frac{-w_0}{w_1} = -5.88 \rightarrow w_0 = 5.88w_1 \end{cases}$$

Assume $w_0 = 5.88 \rightarrow w_1 = 1, w_2 = -2.35$

پاسخ

$$P_1 = \begin{bmatrix} -1.5 \\ -1.7 \end{bmatrix} P_2 = \begin{bmatrix} 0 \\ 2.8 \end{bmatrix} \rightarrow m = \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1} = \frac{0 + 1.5}{2.8 + 1.7} = \frac{1.5}{4.5} = \boxed{0.33}$$

$$y - 2.8 = 0.33(x - 0) \rightarrow \boxed{y = 0.33x + 2.8}$$

$$\rightarrow x_1 = \frac{-w_2}{w_1}x_2 - \frac{w_0}{w_1} \rightarrow \begin{cases} \frac{-w_2}{w_1} = 0.33 \rightarrow w_2 = -0.33w_1 \\ \frac{-w_0}{w_1} = 2.8 \rightarrow w_0 = -2.8w_1 \end{cases}$$

$$\text{Assume } w_0 = 2.8 \rightarrow w_1 = -1, w_2 = 0.33$$

۳. شگرد هسته^۱ چیست و چگونه می‌توان قسمت قبل را با آن حل نمود؟ توضیح دهید.

^۱Kernel trick