به نام خدا

دانشگاه صنعتی خواجه نصیرالدین طوسی

دانشکده برق

درس مبانی سیستم‌های هوشمند

مینی پروژه2

نام: سید محمد رضا حسینی

شماره دانشجویی:40117463

لینک گیت هاب:

https://github.com/isrezahosseini-cell/Fundamentals-of-Intelligent-Systems/tree/main

پرسش اول)

1. فرم اولیه SVM سخت حاشیه(Primal Form)

هدف: یافتن یک ابر صحه جداکننده که کلاس‌ها را بطور کامل جداکند و بیشترین حاشیه‌را **داشته باشد.**

$$Minimize: \frac{1}{2}\|w\|^2$$

تحت محدودیت(constraint):

$$y_i(w.x_i + b) \geq 1, \qquad for\ i = 1, \dots, n$$

2. اثبات عرض حاشیه هندسی

عرض جاشیه هندسی(Geometric Margin) برابر با $\frac{2}{\|w\|}$ است.

مراحل اثبات

1. خطوط حاشیه: ابرصفحه جداکننده در وسط دو خط حاشیه موازی قرار دارد که بردارهای پشتیبان(نزدیکترین داده‌ها) روی ان‌ها قرار می‌گیرند. معادلات این خطوط عبارتند از:

   $w.x_+ + b = +1$ (برای بردارهای پشتیبان با $y_i = +1$)

   $w.x_+ + b = -1$ (برای بردارهای پشتیبان با $y_i = -1$)

2. اختلاف برداری: معادلات را از هم کم می‌کنیم(x_+ و x_- بردارهای پشتیبان هستند):

$$(w.x_+) - (w.x_-) = 1 - (-1)$$

$$w.(x_+ - x_-) = 2$$

3. محاسبه فاصله کل(عرض حاشیه): عرض حاشیه d برابر با طول بردار اختلاف$(x_+ - x_-)$ بر روی بردار واحد عمود بر خطوط، یعنی $\frac{w}{\|w\|}$، است.

$$d = \frac{|w.(x_+ - x_-)|}{\|w\|}$$

4. نتیجه نهایی: با چایگزینی $w.(x_+ - x_-) = 2$، به نتیجه می‌رسیم:

$$\text{عرض حاشیه} = d = \frac{2}{\|w\|}$$

3. استخراج فرم دوگانه و توجیه بردارهای پشتیبان

الف. استخراج فرم دوگانه(Dual From)

برای استخراج فرم دوگانه، از ضرب‌کننده‌های لاگرانژ($a_i \geq 0$) استفاده می‌کنیم.

1. تابع لاگرانژ(Lagrangian):

$$\mathcal{L}(w,b,a) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} a_i[y_i(w.x_i + b) - 1]$$

2. شرط بهینگی(KKT): برای یافتن نقطه زینی(Saddle Point)، $\mathcal{L}$ را نسبت به wو b مشتق می‌گیریم و بابر صفر قرار می‌دهیم:

مشتق نسبت به w:

$$\frac{\partial \mathcal{L}}{\partial w} = -\sum_{i=1}^{n} a_i y_i x_i = 0 \Rightarrow w = \sum_{i=1}^{n} a_i y_i x_i$$

مشتق نسبت به W:

$$\frac{\partial \mathcal{L}}{\partial b} = -\sum_{i=1}^{n} a_i y_i = 0 \Rightarrow \sum_{i=1}^{n} a_i y_i = 0$$

3. جایگذاری(Dual Formation): مقادیر w و شرط مشتق b را در تابع لاگرانژ اولیه جایگذاری می‌کنیم. پس از ساده‌سازی، به فرم دوگانه می‌رسیم:

$$Maximize: \tilde{\mathcal{L}} = (a) = \sum_{i=1}^{n} a_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j y_i y_j (x_i . x_j)$$

تحت محدودیت:

$$\sum_{i=1}^{n} a_i y_i = 0 \ \ and \ a_i \geq 0, for \ i = 1, \dots, n$$

ب. دلیل ظهور بردارهای پشتیبان در تابع تصمیم

تابع تصمیم(Decision Function) SVM برای یک نقطه جدی x_{new} به شکل زیر است:

$$f(x_{new}) = w. x_{new} + b$$

با جایگذاری $w = \sum_{i=1}^{n} a_i y_i x_i$ در تابع تصمیم:

$$f(x_{new}) = \left(\sum_{i=1}^{n} a_i y_i x_i\right). x_{new} + b = \sum_{i=1}^{n} a_i y_i \,(x_i . x_{new}) + b$$

دلیل اینکه نمونه‌ها با $a_i > 0$(بردارهای پشتیبان) در تابع تصمیم ظاهر می‌شوند، شرط مکمل بودن (Slackness Complementary) از شرایط KKTاست:

$$a_i[y_i(w. x_i + b) - 1] = 0$$

این شرط به ما می‌گوید که برای هر نقطه داده i، دو حالت داریم:

1. نقاط غیرپشتیبان: اگر نقطه‌ای خارج از حاشیه باشد، محدودیت فعال نیست:$y_i(w.x_i + b) > 1$.
   برای اینکه شرط KKT برقرار باشد، باید$a_i = 0$ باشد.
2. بردارهای پشتیبان: اگر نقطه‌ای روی مرز حاشیه باشد، محدودیت فعال است:$y_i(w.x_i + b) = 1$.
   در این حالت، $a_i > 0$ خواهد بود.

از انجاییی که در فرم نهایی تابع تصمیم، $f(x_{new})$ یک مجموع وزن‌دار است که با a_i ضرب شده است، فقط نمونه هایی که $a_i > 0$ دارند(یعنی بردارهای پشتیبان) در محاسبه نهایی تابع تصمیم مشارکت می‌کنند و سایر نقاط چون$a_i = 0$ دارند، از این مجموع حذف میشوند.

پرسش دوم)

1. فرم اولیه SVM حاشیه نرم با متغیرهای شل‌شدگی(ξ_i)

فرم اولیه SVM حاشیه نرم برای مدل‌های خطی غیرقابل تفکیک خطی، با اضافه کردن متغیرهای شل‌شدگی(Slack Variables)، $\xi_i \geq 0$ و پارامتر جریمه C، به صورت زیر تعریف می‌شوند.

هدف(Objective): کمینه‌سازی نرم بردار وزن به همراه جریمه خطی برای هرگونه نقض حاشیه.

$$Minimize: \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i$$

تحت محدودیت (Constraints):

1. محدودیت شل‌شدگی(Margin Constraint with Slack):

$$y_i(w.x_i + b) \geq 1 - \xi_i, \qquad for\ i = 1, \dots, n$$

2. نامنفی بودن متغیرهای شل‌شدگی:

$$\xi_i \geq 1, \qquad for\ i = 1, \dots, n$$

نقش متغیرهای ξ_i

$\xi_i = 0$: داده به درستی طبقه‌بندی شده و خارج یا روی حاشیه قرار دارد(مانندHard Margin SVM).

$0 < \xi_i < 1$: داده به درستی طبقه‌بندی شده، اما در داخل حاشیه قرار گرفته است(نقض نرم حاشیه).

$\xi_i \geq 1$: داده به اشتباه طبقه‌بندی شده است(در سمت اشتباه ابرصفحه قرار دارد).

2.اثبات تساوی با تابع زیان لولا(Hinge Loss)

فرمول اولیه را می‌توان با کمینه‌سازی تابع زیر دانست، که در ان زیان لولا(Hinge Loss)استفاده شده است.

$$Minimize\colon \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \max\,(0, 1 - y_i f(x_i))$$

که در ان $f(x_i) = w.x_i + b$.

مراحل اثبات:

هدف ما ثابت کردن این است که:

$$Minimize\left[\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi_i\right]$$

تحت محدودیت $\xi_i \geq 1 - y_i f(x_i)$ و$\xi_i \geq 0$، معادل است با:

$$Minimize\ \left[\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \max\bigl(0, 1 - y_i f(x_i)\bigr)\right]$$

1. تعریف ξ_iبهینه: متغیرهای شل‌شدگی ξ_i در تابع هدف اولیه(بخش 1)دارای ضریب مثبت C هستند. بنابراین برای کمینه‌سازی کل عبارت، باید مقادیر ξ_i را تا حد امکان کوچک انتخاب کنیم.
2. شرط بر روی ξ_i: متغیر ξ_i: باید همزمان دو شرط را ارضا کند:

شرط اول(از محدودیت): $\xi_i \geq 1 - y_i f(x_i)$

شرط دوم(از تعریف): $\xi_i \geq 0$

3. انتخاب مقدار کمینه: برای کمینه‌سازی تابع هدف، باید ξ_i را برابر با بیشینه (Maximum) دو مقدار کران پایین قرار دهیم:

$$\xi_i^* = \max\,(0, 1 - y_i f(x_i))$$

هرمقدار ξ_i کوچکتر از ξ_i^*، حداقل یکی از محدودیت‌ها را نقض می‌کند.

4. چایگزینی: با جایگزینی مقدار بهینه ξ_i^* در تابع هدف اولیه:

$$Minimize \left[\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi_i^*\right]$$

$$Minimize \; \left[\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \max\big(0, 1 - y_i f(x_i)\big)\right]$$

این عبارت $\max\,(0, 1 - y_i f(x_i))$ همان زیان لولا(Hinge Loss) است.

3. تاثیر پارامتر C بر تعمیم‌پذیری و بردارهای پشتیبان

پارامتر C یک پارامتر تنظیم‌کننده(Regularization Parameter) است که میزان تعادل بین دو هدف متضاد در SVM حاشیه نرم کنترل می‌کند:

بیشینه سازی حاشیه: که با کمینه سازی $\frac{1}{2}\|w\|^2$ کنترل می‌شود.

کمینه سازی خطای طبقه بندی و نقض حاشیه: که با کمینه سازی $\sum \xi_i$ (زیان لولا) کنترل می‌شود.

الف. تاثیر بر تعمیم2پذیری(Generalization)

C کوچک:

$\frac{1}{2}\|w\|^2$(جریمه حاشیه) اهمیت بیشتری داده می‌شود.

SVM ترجیح می‌دهد حاشیه بزرگتری را انتخاب کند، ختی اگر تعداد بیشتری از نقاط در داخل حاشیه قرار گیرند یا اشتباه طبقه‌بندی شوند(خطای اموزش افزایش یابد).

این مدل ساده‌تر است(نرمw کوچکتر است) و احتمال بیش برازش(Overfitting) کاهش میابد.

بنابراین، تعمیم‌پذیری به طور کلی بهتر است.

C بزرگ:

به $\sum \xi_i$(جریمه خطا) اهمیت بیشتری داده می‌شود.

SVM تلاش می‌کند تمام نقاط را به درستی طبقه‌بندی کرده و نقض حاشییه را به حداقل برساند (خطای اموزش کمینه می‌شود).

این مدل ممکن است پیچیده‌تر شود(نرم w بزرگتر است) و احتمال بیش برازش افزایش می‌یابد.

تعمیم‌پذیری ممکن است کاهش یابد.

ب. تاثیر بر تعداد بردارهای پشتیبان

C کوچک:

مدل انعطاف پذیری بیشتری در نقض حاشیه دارد. بسیاری از نقاط ممکن است در داخل حاشیه قرار گیرند($0 < \xi_i < 1$) یا حتی اشتباه طبقه بندی شوند($\xi_i \geq 1$).

هر نقطه‌ای که $\xi_i > 0$ یا دقیقا روی مرز حاشیه باشد، بردار پشتیبان محسوب می‌شود.

در نتیجه بردارهای پشتیبان افزایش می‌یابد.

C بزرگ:

مدل جریمه سنگینی را برای نقض حاشیه اعمال می‌کند. فقط نقاطی که عملا جدایی‌ناپذیرند یا نزدیکترین نقاط به مرزهای سخت حاشیه هستند، به عنوان بردار پشتیبان باقی می‌مانند.

مدل تلاش می‌کند تا حد ممکنبه SVM سخت حاشیه نزدیک شود.

در نتیجه، تعداد بردارهای پشتیبان کاهش می‌یابد.(التبه در حالت حدی که $C \longrightarrow \infty$ می‌شود ، به مدل سخت حاشیه باز می‌گردیم که تعداد بردارهای پشتیبان می‌تواند زیاد باشد).

به طور خلاصه، C نقش یک پارامتر معاوضه(Trade-off)را دارد: C بزرگ یعنی تاکیید بر دقت اموزش(مدل پیچیده)، و C کوچک یعنی تاکیید بر تعمیم‌پذیری(مدل ساده‌تر).

پرسش سوم)

الف. فرم اولیه Soft Margin-ℓ_2 SVM

در این مدل، به جای جریمه خطی(نرم ℓ_1) برای متغیرهای شل‌شدگی ξ_i، از جریمه مربعی(ℓ_2) استفاده می‌شود.

$$Minimize\left[\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i^2\right]$$

تحت محدودیت(Constraints):

$$y_i(w.x_i + b) \geq 1 - \xi_i, \qquad for\ i = 1, \ldots, n$$

توجه: برخلاف حالت ℓ_1، محدودیت $\xi_i \geq 0$ دیگر ضروری نیست، زیرا جریمه ξ_i^2 همیشه مثبت است و کمینه سازی تضمین می‌کند که ξ_iها تا حد امکان کوچک و نزدیک به صفر باشند.

ب. دوگان متناظر(Dual Form)

برای استخراج دوگان، از تابع لاگرانژ استفاده می‌کنیم:

$$\mathcal{L}(w, b, \xi, a) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i^2 - \sum_{i=1}^{n}a_i[y_i(w.x_i + b) - 1 + \xi_i]$$

مشتق نسبت به w و b همانند حالت ℓ_1 است:

$$\frac{\sigma\mathcal{L}}{\sigma w} = 0 \Longrightarrow w = \sum\nolimits_{i=1}^{n} a_i y_i x_i$$

$$\frac{\sigma\mathcal{L}}{\sigma b} = 0 \Longrightarrow \sum\nolimits_{i=1}^{n} a_i y_i = 0$$

مشتق نسبت به ξ_i: این بخش، تفاوت اصلی را ایجاد می‌کند.

$$\frac{\sigma \mathcal{L}}{\sigma\ \xi_i} = 2C\ \xi_i - a_i = 0 \Longrightarrow \xi_i = \frac{a_i}{2C}$$

با جایگذاریb،w و ξ_i در$\mathcal{L}$ ، فرم دوگانه به دست می‌اید:

$$Maximize{:}\ \tilde{\mathcal{L}}(a) = \sum_{i=1}^{n} a_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j y_i y_j (x_i . x_j) - \sum_{i=1}^{n} \frac{a_i^2}{4C}$$

تحت محدودیت:

$$\sum_{i=1}^{n} a_i y_i = 0\ and\ a_i \geq 0, \qquad for\ i = 1, \ldots, n$$

ج. تفاوت کران‌های a_i با حالت ℓ_1 و اثر ان بر SVها

ویژگی	ℓ_1Soft Margin(جریمه خطی)	ℓ_1Soft Margin(جریمه مربعی)
کران‌های a_i در دوگان	$0 \leq a_i \leq C$	$0 \leq a_i$ (بدون کران بالایی)
تعداد بردارهای پشتیبان(SV)	a_i ها بهC محدود می‌شوند، که اغلب منجر به تعداد کمتری SVمی‌شود.	a_i ها کران بالا ندارند، که اغلب منجر به تعداد بیشتری SV(وSVهای با a_iکوچک)می‌شود.
وزنSVهای خارج حاشیه	SVهای خارج حاشیه($\xi_i \leq 1$) همیشه= a_i C دارند.(نقاط پرت به شدت جریمه و سخت محدود می‌شوند)	SVهای خارج حاشیه($\xi_i \geq 1$)می‌توانند $a_i > C$ داشته باشند.(نقاط پرت به صورت نرم تر و پیوسته‌تر جریمه می‌شوند.)

توضیح اثر بر w (وزن‌ها)

در هر دو حالت، $w = \sum_{j=1}^{n} a_i y_i x_i$.

ℓ_1: دارای کران سخت $a_i \leq C$ است. این امر باعث می‌شود که یک نقض حاشیه بزرگ(نقاط پرت) فورا a_i را به C برساند و سهم ان در w محدود باشد. این مدل کمی تنک تر(Sparser) است.

ℓ_2:کران a_iندارد. $\xi_i = \frac{a_i}{2C}$. این بدان معناست که a_i می‌تواند بزرگ باشد تا به یک نقطه پرت بزرگ پاسخ دهد. این جریمه پیوسته باعث می‌شود که $\ell_2 - SVM$نسبت به نقاط پرت حساس‌تر باشد و ممکن استw بزرگتر و متراکم‌تری تولید کند.

پرسش چهارم) کرنل و شرایط KKT

الف. تعریف کرنل معتبر و ارتباط با نگاشت ویژگی

کرنل معتبر(Valid Kernel)، که با K(x_i،x_j) نشان داده می‌شود، یک تابع ریاضی است که شرایط خاصی را ارضا می‌کند تا بتواند معادل با ضرب داخلی(Dot Product) در یک فضای ویژگی با بعد بالاتر(احتمالا بی نهایت)باشد.

$$K(x_i, x_j) = \phi(x_i).\phi(x_j)$$

$\phi(.)$ نگاشت ویژگی (Feature Mapping) است مه بردار ورودیx را از فضای ورودی بعد پایین به یک فضای ویژگی بعد بالاتر $\mathcal{H}$ نگاشت می‌کند: $\phi: \mathcal{X} \longrightarrow \mathcal{H}$.

شرط اصلی معتبر بودن(قضیه مرسر-Mercer’s Theorem):یک تابعK زمانی یک کرنل معتبر است که ماتریس کرنلK($K(x_i, x_j) = K_{ij}$)برای هر مجموعه محدودی از نقاط ورودی، یک ماتریس مثبت نیمه معین(Positive Semi-Definite Matrix)باشد.

ارتباط با نگاشت ویژگی: کرنل معتبر به ما اجازه می‌دهد بدون نیاز به تعریف و محاسبه صریح$\phi(x)$ (که اغلب بسیار پیچیده یا بی نهایت بعد است)، ضرب داخلی$\phi(x_i).\phi(x_j)$ را در فضای بعد بالا محاسبه کنیم. این تکنیک به ترفند کرنل($Kernel\ Trick$) معروف است.

ب. خواص کرنل های RBF و چند جمله‌ای

کرنل‌هایRBF و چند جمله‌ای دو مثال رایج از کرنل معتبر هستند:

کرنل چند جمله‌ای(Polynomial Kernel):

$$K(x_i, x_j) = (x_i.x_j + c)^d$$

خواص بردار(Power Property): اگر k_1و k_2 کرنل باشند، $K(x_i, x_j) = K_1(x_i.x_j)^d$ نیز کرنل معتبر است($d\epsilon\mathbb{N}$). این خاصیت تضمین میکند که توانd ام از یک کرنل معتبر(مانند ضرب داخلی ساده)، همچنان یک کرنل معتبر خواهد بود. این کرنل یک نگاشت به یک فضای متناهی را انجام می‌دهد.

کرنل تابع شعاعی(RBF-Radial Basis Function/Gaussian Kernel):

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$$

خواص جمع و ضرب(Summation and Scaling Property): اگر K_1 کرنل باشد، $\exp(K_1)$ نیز کرنل معتبر است.

کرنل RBF بر اساس قضیه تیلور(Taylor Expansion) می‌تواند به عنوان یک ضرب داخلی در یک فضای ویژگی بعد نامتناهی(Infinite-Dimensional) در نظر گرفته شود. این خاصیت به ان اجازه می‌دهد تا مرزهای تصمیم بسیار پیچیده‌ای را مدل کند.

ج. حالاتKKT به صورت مرتب و نقش هر حالت

شرایط کروش-کوهن-تاکر(Karush-Kuhn-Tucher-KKT)مجموعه‌ای از شرایط لازم بهینگی(Optimality) هستند که باید توسط هر جواب بهینه(w^*, b^*) برای یک مسئله بهینه‌سازی مقید محدب (مانندSVM) ارضا شوند.

حالاتKKT (برای SVM حاشیه نرم ℓ_1)

حالت	شرطKKT	توصیف هندسی و نقش
مشتق w	$w = \sum_{i=1}^{n} a_i y_i x_i$	بردار وزن‌ها را به صورت ترکیب خطی از نمونه‌ها و ضرب کننده‌های لاگرانژ بیان می‌کند.
مشتق b	$\sum_{i=1}^{n} a_i y_i = 0$	شرط تعادل و تقارن لازم برای قرارگیری ابرصفحه در فضای ویژگی.
مکمل بودن	$a_i[y_i f(x_i) - 1 + \xi_i] = 0$	تعیین‌کننده نوع داده: اگر$a_i > 0$ باشد، محدودیت فعال است(داده روی مرز یا داخل ان است).
دوگان‌پذیری	$0 \le a_i \le C$	کران‌های ضرب‌کننده‌های لاگرانژ(تنظیم کننده جریمه)
اولیه/ شل شدگی	$y_i f(x_i) \ge 1 - \xi_i, \xi_i \ge 0$	تمام نقاط باید به درستی طبقه‌بندی شوند(با لحاظ شل شدگی)

نقش هر حالت:

شرایط KKT نقش حیاتی در تشخیص بردارهای پشتیبان و اجرای الگوریتم‌های SVM دارند:

نقاط غیرپشتیبان(Non-SV):

$y_i f(x_i) = 1$ (خارج از حاشیه) $\xi_i = 0$.

از شرط مکمل بودن، چون عبارت داخل براکت مثبت است، باید $a_i = 0$باشد. این نقاط هیچ تاثیری بر w ندارند.

بردارهای پشتیبان معمولی(SV on Margin):

$y_i f(x_i) = 1$(دقیقا روی مرز)$\xi_i = 0$.

از شرط مکمل بودن، عبارت داخل براکت صفر است. $0 < a_i < C$.

بردارهای پشتیبان خطا(SV violating Margin):

$y_i f(x_i) < 1$(نقض حاشیه یا اشتباه طبقه‌بندی).

از شرط مکمل بودن، a_i باید C باشد. این‌ها بحرانی ترین نقاط هستند.

پرسش پنجم)

داده ها

کلاس مثبت($y_i = +1$):(2, 0), (3, 0), (2, 1), (2, -1)

کلاس منفی($y_i = -1$):(-2, 0), (-3, 0), (-2, 1), (-2, -1)

الف. حدس ابر صفحه حداکثر حاشیه و معادله مرز تصمیم

حدس ابرصفحه

با مشاهده نقاط داده، متوجه می‌شویم که داده‌ها حول محورy به طور کامل متقارن هستند.

تمام نقاط مثبت در ناحیه$x > 0$ و تمام نقاط منفی در ناحیه $x < 0$ قرار دارند.

فاصله نقاط مثبت از محور y، حداقل 2 واحد است.

فاصله نقاط منفی از محورy، حداقل 2 واحد است.

بنابراین، بهترین و متقارن‌ترین ابرصفحه جداکننده(خط جداکننده در $\mathbb{R}^2$)باید محورy باشد.

معادله مرز تصمیم(ابرصفحه)

معادله محورy به صورت زیر است:

$$x = 0 \text{ یا } 1.x + 0.y = 0$$

معادله مرز تصمیم:

$$w^\intercal x + b = 0 \implies x = 0$$

محاسبه دستی ضرایب w و b

معادله ابرصفحه تصمیم ما x=0 است، که به این معنی است:

$$w = (w_1, w_2) \text{ و } x = (x_1, x_2)$$

$$w_1 x_1 + w_2 x_2 + b = 0 \Rightarrow 1. x_1 + 0. x_2 + 0 = 0$$

پس، ما حدس می‌زنیم:

$$w \propto (1,0) \text{ و } b = 0$$

حالا باید مقادیر دقیق w و b را طوری تنظیم کنیم که محدودیت سخت حاشیه را برای بردارهای پشتیبان(نقاط روی مرز ± 1) ارضا کنند:

$$y_i (w. x_i + b) = 1$$

شناسایی بردارهای پشتیبان

نزدیکترین نقاط به خط $x=0$ نقاطی هستند که $|x_1|$کمینه را دارند. این نقاط عبارتند از:

کلاس مثبت: $x_{SV}^{+} = (2,0), (2,1), (2,-1)$

کلاس منفی: $x_{SV}^{-} = (-2,0), (-2,1), (-2,-1)$

اعمال محدودیت

برای کلاس مثبت $(2,0)$: $y_i = 1, x_i = (2,0)$

$$1. (w_1(2) + w_2(0) + b) = 1 \Longrightarrow 2w_1 + b = 1 (*)$$

برای کلاس منفی $(-2,0)$: $y_i = 1, x_i = (-2,0)$

$$-1.(w_1(-2)+w_2(0)+b)=1 \Longrightarrow 2w_1-b=1(**)$$

حل دستگاه معادلات

با جمع(❋) و (❋❋):

$$(2w_1+b)+(2w_1-b)=1+1$$

$$4w_1=2 \Rightarrow w_1=\frac{1}{2}$$

با قرا ر دادن$w_1=\frac{1}{2}$در (❋):

$$2\left(\frac{1}{2}\right)+b=1 \Rightarrow 1+b=1 \Rightarrow b=0$$

همچنین، چون در حدس اولیه$w \propto (1,0)$داشتیم،w_2باید صفر باشد تا ابرصفحه به مختصهx_2(محور y) وابسته نباشد:

$$w_2=0$$

ضرایب محاسبه شده:

$$w=\left(\frac{1}{2},0\right), b=0$$

ج. بردارهای پشتیبان، معادلات ابرصفحه±1 و مقدار حاشیه هندسی

بردارهای پشتیبان(Support Vectors-SVs)

بردارهای پشتیبان، تمام نقاطی هستند که در محاسبه w و b محدودیت را فعال می‌کنند(یعنی روی مرز ± 1 قرار دارند):

$$SV=\{(2,0),(2,1),(2,-1),(-2,0),(-2,1),(-2,-1)\}$$

معادلات ابرصفحه $w^\intercal x + b = \pm 1$

با استفاده از $w = (\frac{1}{2}, 0)$ و $b = 0$:

$$w^\intercal x + b = \pm 1 \Rightarrow \frac{1}{2}x_1 + 0x_2 + 0 = \pm 1$$

ابرصفحه حاشیه مثبت(+1):

$$\frac{1}{2}x_1 = 1 \Rightarrow x_1 = 2$$

ابرصفحه حاشیه مثبت(-1):

$$\frac{1}{2}x_1 = -1 \Rightarrow x_1 = -2$$

مقدار حاشیه هندسی(Geometric Margin)

حاشیه هندسی برابر است با $\frac{1}{\|w\|}$، که فاصله ابر صفحه تصمیم تا نزدیک‌ترین نقطه است.

$$\|w\| = \sqrt{(\frac{1}{2})^2 + 0^2} = \sqrt{\frac{1}{4}} = \frac{1}{2}$$

فاصله تا ابرصفحه تصمیم(حاشیه هندسی):

$$\text{حاشیه هندسی} = \frac{1}{\|w\|} = \frac{1}{1/2} = 2$$

فاصله بین دو ابرصفحه($\frac{2}{\|w\|}$):

$$\text{عرض حاشیه} = \frac{2}{\|w\|} = \frac{2}{1/2} = 4$$

این مقدار منطقی است، زیرا فاصله بین خط $x_1 = 2$ و $x_1 = -2$ دقیقا 4 واحد است.

پرسش ششم)

الف. پیشنهاد مدل مناسب با استفاده از شبکه‌های عصبی

هدف ما مدل‌سازی مرزهای تصمیم پیچیده و غیرخطی است. برای این منظور، یک شبکه عصبی چندلایه(Multi-Layer-Perceptron MLP) بهترین گزینه را ارائه می‌دهد.

مدل پیشنهادی:

شبکه عصبی چندلایه(MLP): MLP شامل یک لایه ورودی، یک یا چند لایه مخفی، و یک لایه خروجی است.

ساختار پیشنهادی(به عنوان مثال):

لایه ورودی: 2 گره(به ازای مختصات xو y).

لایه مخفی: حداقل دو لایه مخفی با تعداد گره‌های کافی(مثلا 10 گره در هر لایه) برای یدگیری ویژگی‌های غیر خطی.

توابع فعال‌ساز(Activation Functions) لایه‌های مخفی: برای معرفی خاصیت غیرخطی، باید از توابع فعال‌ساز غیرخطی مانند(Rectified Linear Unit) ReLU یاSigmoid/tanh استفاده شود.

لایه خروجی: 1 گره(با تابع فعال‌ساز سیگموئید برای طبقه‌بندی دوگانه).

توابع پیشنهادی:

Sigmoid یاRelU :

Sigmoid: در گذشته رایج بود، اما به دلیل مشکل شیب محوشونده(Vanishing Gradient)در لایه‌های عمیق کمتر استفاده می‌شود.

ReLU: به دلیل سادگی و جلوگیری از مشکل شیب محوشونده، معمولا انتخاب ارجح در لایه‌های میانی است.

نتیجه: برای طبقه‌بندی این داده‌های غیرخطی، یک MLP با توابع فعال‌سازReLU در لایه‌های مخفی مناسب‌ترین انتخاب است.

ب. ضعف سایر مدل‌های شبکه عصبی و عملکرد ضعیف مدل‌های خطی

1. ضعف MLP با توابع فعال‌ساز سراسری(Sigmoid/tanh)

اگرچهMLP به خودی خود می‌تواند مرزهای غیرخطی را مدل کند، استفاده از توابعی مانندSigmoidیا tanh در لایه‌های مخفی ممکن است در این مسئله خاص(داده‌های خوشه‌ای) به طور موثر عمل نکند.

توابع فعال ساز سراسری(Global Activation Functions): توابعی مانندSigmoidو tanh شیب‌های خود را در نواحی دور از صفر از دست می‌دهند(شیب محوشونده). در یک معماری عمیق، این امر باعث می‌شود یادگیری ویژگی‌ها کند شود.

مشکل کرزهای تصمیم غیر محلی(Non-local Decision Boundaries):توابع فعال‌ساز سراسری به طور ذاتی تمایل به ایجاد مرزعای تصمیم صاف و سراسری دارند. این مسئله باعث می‌شود مدل برای جداکردن چهار ناحیه خوشه‌ای کاملا مجزا(مانند شکل 5 در سوالات) دچار مشکل شود و ممکن است مرزهای تصمیم، بیش از حد از نقاط اموزش دور شوند. در مقابل، توابعی مانندReLU یا مدل‌های بر پایه کرنل بهتر می‌توانند رفتار موضعی(Local)داده را حفظ کنند.

2.عملکرد ضعیف مدل‌های خطی در تفکیک خوشه‌های محلی

مدل‌های خطی(مانندSVMخطی یا MLP بدون لایه‌های مخفی غیر خطی) عملکرد بسیار ضعیفی در تفکیک این داده‌ها دارند زیرا:

فرض تفکیک‌پذیری خطی: مدل‌های خطی فرض می‌کنند که می‌توان دو کلاس را با یک خط مستقیم(یا ابر صفحه در ابعاد بالاتر) از هم جدا کرد.

ناتوانی در مدیریت XOR : این مجموعه داده، یک نوع تعمیمیم‌یافته از مسئله (XOR (Exclusive OR در فضای بعد پایین است. برای طبقه‌بندی یک نقطه جدید در مرکز، خط جداکننده مجبور است یک کلاس (مثلا ابی) را در دو ناحیه جدا از هم قرار دهد، که توسط یک خط مستقیم امکان‌پذیر نیست.

نیاز به نگاشت به فضای بالاتر: برای تفکیک خطی این خوشه‌ها، داده‌ها ابتدا باید توسط یک نگاشت ویژگی غیرخطی(مانند نگاشت$\phi(x)$ در SVM کرنل یا همان لایه‌های مخفی در MLP) به یک فضای بعد بالاتر نگاشته شوند تا در انجا به صورت خطی قابل تفکیک باشند. مدل های خطی این توانایی را ندارند.

پرسش پیاده سازی RBF)

هدف این پرسش، مدل‌سازی یک سیستم دینامیکی ساده با استفاده از یک شبکه عصبی با تابع شعاعی پایه است. در گام اول با پیاده‌سازی یک RBFNN استاندارد با ساختاری ایستا شروع می‌شود (بخش اول) و سپس یک شبکه پیشرفته‌تر و تطبیقی(بخش دوم) توسعه داده می‌شود که ساختار خود را بر اساس شبکه M-RAN که در مقاله ارائه شده، بهینه سازی می‌شود.

قسمت اول:

شناسایی سیستم با شبکه عصبی RBF

برای این سوال ابتدا یک سیستم ساده شده از نوع ball and beam را مدل‌سازی می‌شود. موقعیت توپ در زمان**t**، که با y(t)نشان داده می‌شود، تابعی از دو موقعیت قبلی خود یعنیy(t-1)و y(t-2) و ورودی قبلی به سیستم یعنی u(t-1) است. این رابطه توسط معادله تفاضلی زیر تعریف می‌شود:

$$y(t) = \frac{y(t-1)}{1+y(t-2)} + u(t-1)^3$$

گام نخست: ایجاد دیتا ست

📝 پاسخ به بخش ۳.۱.۱: تحلیل مسئله

۱. چرا این مسئله یک مسئله تقریب تابع است؟

- **تعریف مسئله:** مسئله‌ای که در اینجا تعریف شده، پیش‌بینی مقدار خروجی $y(t)$ بر اساس یک بردار ورودی **معین** است: $\mathbf{x}(t) = [y(t-1), y(t-2), u(t-1)]^\top$.
- **وابستگی به فرمول:** اگرچه سیستم توسط یک معادله تفاضلی جبری (فرمول ۶۲) تعریف شده است، در حقیقت، **شبکه عصبی به فرمول دسترسی ندارد.**
- **وظیفهٔ شبکه:** وظیفه شبکه عصبی RBFNN این است که **یک نگاشت (Mapping)** بین فضای ورودی سه‌بعدی ($\mathbf{x}(t)$) و فضای خروجی تک‌بعدی ($y(t)$) را **تخمین** بزند.
- **نتیجه:** در نهایت، شبکه تلاش می‌کند تا یک تابع **ناشناخته (مجهول)** f را بیابد به طوری که $y(t) \approx f(\mathbf{x}(t))$ باشد. از آنجا که هدف تخمین (تقریب) یک تابع $f : \mathbb{R}^3 \to \mathbb{R}$ از طریق داده‌های ورودی-خروجی است، این یک **مسئله تقریب تابع** نامیده می‌شود.