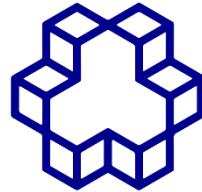


به نام خدا



دانشگاه صنعتی خواجه نصیرالدین طوسی

دانشکده برق

درس مبانی سیستم‌های هوشمند

میزی پروژه ۲

نام: سید محمد رضا حسینی

شماره دانشجویی: **40117463**

نام : علی کاشی پزها

شماره دانشجویی: **40121723**

لینک گیت هاب:

[https://github.com/isrezahosseini-cell/Fundamentals-of-
Intelligent-Systems/tree/main](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 صفحه جداکننده که کلاس‌ها را بطور کامل جداکند و بیشترین حاشیه را داشته باشد.

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

تحت محدودیت (constraint):

$$y_i(w \cdot x_i + b) \geq 1, \quad \text{for } i = 1, \dots, n$$

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 آن‌ها قرار می‌گیرند. معادلات این خطوط عبارتند از:

$$(y_i = +1) w \cdot x_+ + b = +1$$

$$(y_i = -1) w \cdot x_+ + b = -1$$

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

$$(w \cdot x_+) - (w \cdot x_-) = 1 - (-1)$$

$$w \cdot (x_+ - x_-) = 2$$

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.

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

4. نتیجه نهایی: با چایگرینی $2 \cdot w \cdot (x_+ - x_-) = 2$ به نتیجه می رسیم:

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

3. استخراج فرم دوگانه و توجیه بردارهای پشتیبان

الف. استخراج فرم دوگانه (Dual From)

برای استخراج فرم دوگانه، از ضرب کننده‌های لاغرانژ ($a_i \geq 0$) استفاده می کنیم.

1. تابع لاغرانژ (Lagrangian):

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

2. شرط بهینگی (KKT): برای یافتن نقطه زینی (\mathcal{L} , Saddle Point) را نسبت به w و b مشتق می گیریم

و با بر صفر قرار می دهیم:

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 b :

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

3. جایگذاری (Dual Formation): مقادیر w و شرط مشتق b را در تابع لAGRANZ اولیه جایگذاری

می‌کنیم. پس از ساده‌سازی، به فرم دوگانه می‌رسیم:

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

تحت محدودیت:

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

ب. دلیل ظهور بردارهای پشتیبان در تابع تصمیم

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

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

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

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

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

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

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: $y_i(w \cdot x_i + b) > 1$.

برای اینکه شرط KKT برقرار باشد، باید $a_i = 0$ باشد.

2. بردارهای پشتیبان: اگر نقطه‌ای روی مرز حاشیه باشد، محدودیت فعال است: $y_i(w \cdot x_i + b) = 1$.

در این حالت، $a_i > 0$ خواهد بود.

از انجایی که در فرم نهایی تابع تصمیم، $f(x_{new})$ یک مجموع وزن دار 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 شل شدگی(ξ_i)

فرم اولیه SVM حاشیه نرم برای مدل‌های خطی غیرقابل تفکیک خطی، با اضافه کردن متغیرهای شل شدگی(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.

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

تحت محدودیت (Constraints)

1. محدودیت شل شدگی(Margin Constraint with Slack)

$$y_i(w \cdot x_i + b) \geq 1 - \xi_i, \quad \text{for } i = 1, \dots, n$$

2. نامنفی بودن متغیرهای شل شدگی:

$$\xi_i \geq 0, \quad \text{for } i = 1, \dots, n$$

نقش متغیرهای ξ_i

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

$\xi_i < 0$: 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.

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

که در ان $f(x_i) = w \cdot x_i + b$

مراحل اثبات:

هدف ما ثابت کردن این است که:

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

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

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

1. تعریف ξ_i : بهینه: متغیرهای شل شدگی ξ_i درتابع 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 ارضاء کند:

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

$$\xi_i \geq 0 \quad \text{شرط دوم (از تعریف):}$$

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

کران پایین قرار دهیم:

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

هر مقدار 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:

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

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

این عبارت (Hinge Loss) همان زیان لولا 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 طبقه‌بندی و نقض حاشیه: که با کمینه سازی $\sum_i \xi_i$ (زیان لولا) کنترل می‌شود.

الف. تاثیر بر تعمیم‌پذیری (Generalization)

C کوچک:

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

SVM ترجیح می‌دهد حاشیه بزرگتری را انتخاب کند، حتی اگر تعداد بیشتری از نقاط در داخل حاشیه

قرار گیرند یا اشتباه طبقه‌بندی شوند (خطای آموزش افزایش یابد).

این مدل ساده‌تر است (نرم w کوچک‌تر است) و احتمال بیش برآش (Overfitting) کاهش می‌ابد.

بنابراین، تعمیم‌پذیری به طور کلی بهتر است.

C بزرگ:

$\sum_i \xi_i$ (جریمه خطأ) اهمیت بیشتری داده می‌شود.

SVM تلاش می‌کند تمام نقاط را به درستی طبقه‌بندی کرده و نقض حاشیه را به حداقل برساند

(خطای آموزش کمینه می‌شود).

این مدل ممکن است پیچیده‌تر شود (نرم w بزرگ‌تر است) و احتمال بیش برآش 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($1 < \xi_i < 0$) یا حتی اشتباه طبقه بنده شوند($\xi_i \geq 1$).

هر نقطه‌ای که $\xi_i < 0$ یا دقیقاً 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 عملاً 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 $C \rightarrow \infty$ 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 SVM

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

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

تحت محدودیت(Constraints):

$$y_i(w \cdot x_i + b) \geq 1 - \xi_i, \quad \text{for } i = 1, \dots, n$$

توجه: برخلاف حالت ℓ_1 ، محدودیت $0 \leq \xi_i \leq \xi_i^2$ دیگر ضروری نیست، زیرا جریمه ξ_i^2 همیشه مثبت است و کمینه

سازی تضمین می‌کند که ξ_i ها تا حد امکان کوچک و نزدیک به صفر باشند.

ب. دوگان متناظر(Dual Form)

برای استخراج دوگان، ازتابع لاغرانژ استفاده می‌کنیم:

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

مشتق نسبت به w و b همانند حالت ℓ_1 است:

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

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

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

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

با جایگذاری w , b , ξ_i در \mathcal{L} ، فرم دوگانه به دست می‌اید:

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

تحت محدودیت:

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

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

ℓ_1 Soft Margin (جریمه مربعی)	ℓ_1 Soft Margin (جریمه خطی)	ویژگی
$0 \leq a_i$ (بدون کران بالایی)	$0 \leq a_i \leq C$	کران‌های a_i در دوگان

a_i ها کران بالا ندارند، که اغلب منجر به تعداد بیشتری SV و SVهای با a_i کوچک) می‌شود.	a_i ها به C محدود می‌شوند، که اغلب منجر به تعداد کمتری SV می‌شود.	تعداد بردارهای پشتیبان (SV)
SVهای خارج حاشیه ($1 \geq \xi_i \leq C$) همیشه $a_i > C$ دارند. (نقاط پرت به شدت جریمه و سخت نرم تر و پیوسته تر جریمه می‌شوند).	$a_i = \frac{a_i}{C}$ دارد. (نقاط پرت به شدت جریمه و سخت محدود می‌شوند)	وزن SVهای خارج حاشیه

توضیح اثر بر w (وزن‌ها)

$$w = \sum_{j=1}^n a_j y_j x_j$$

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

ℓ_2 : کران a_i ندارد. $\xi_i = \frac{a_i}{2C}$. این بدان معناست که a_i می‌تواند بزرگ باشد تا به یک نقطه پرت بزرگ پاسخ دهد. این جریمه پیوسته باعث می‌شود که $SVM - \ell_2$ نسبت به نقاط پرت حساس‌تر باشد و ممکن است w 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 ارضاء می‌کند تا بتواند معادل با ضرب داخلی (Dot Product) در یک فضای ویژگی با بعد بالاتر (احتمالاً بی نهایت) باشد.

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

(.) ϕ نگاشت ویژگی (Feature Mapping) است مه بردار ورودی x را از فضای ورودی بعد پایین به یک

فضای ویژگی بعد بالاتر \mathcal{H} نگاشت می‌کند: $\phi: \mathcal{X} \rightarrow \mathcal{H}$.

شرط اصلی معتبر بودن (قضیه مرسن-Mercer's Theorem): یک تابع K زمانی یک کرنل معتبر است که

ماتریس کرنل $K(x_i, x_j) = K_{ij}$ 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 $(x_i \cdot \phi(x_i)) \cdot (x_j \cdot \phi(x_j))$ را در فضای بعد بالا محاسبه کنیم. این تکنیک به ترفند کرنل (Kernel Trick) معروف است.

ب. خواص کرنل‌های RBF و چند جمله‌ای

کرنل‌های RBF و چند جمله‌ای دو مثال رایج از کرنل معتبر هستند:

: (Polynomial Kernel) کرنل چند جمله‌ای

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

خواص بردار (Power Property): اگر k_1 و k_2 کرنل باشند، $K(x_i, x_j) = K_1(x_i, x_j)^d$ نیز کرنل معتبر است ($d \in \mathbb{N}$). این خاصیت تضمین می‌کند که توان d 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.

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

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

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. حالات KKT 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 (مانند SVM) ارضا شوند.

حالات KKT (برای SVM حاشیه نرم ℓ_1)

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

نقش هر حالت:

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

نقاط غیرپشتیبان (Non-SV):

$$\xi_i = 0 \quad (\text{خارج از حاشیه}) \quad y_i f(x_i) = 1$$

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

بردارهای پشتیبان معمولی (SV on Margin):

$$\xi_i = 0 \quad (\text{دقیقاً روی مرز}) \quad y_i f(x_i) = 1$$

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

بردارهای پشتیبان خطأ (SV violating Margin):

$$y_i f(x_i) < 1 \quad (\text{نقض حاشیه یا اشتباه طبقه‌بندی}).$$

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 محور u 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 u ، حداقل 2 واحد است.

فاصله نقاط منفی از محور u ، حداقل 2 واحد است.

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

معادله مرز تصمیم (ابرصفحه)

معادله محور γ به صورت زیر است:

$$x = 0 \text{ یا } 1 \cdot x + 0 \cdot y = 0$$

معادله مرز تصمیم:

$$w^T x + b = 0 \Rightarrow x = 0$$

محاسبه دستی ضرایب w و b

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

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

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

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) ارضاء کنند:

$$y_i(w \cdot x_i + b) = 1$$

شناسایی بردارهای پشتیبان

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

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

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

اعمال محدودیت

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

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

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

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

حل دستگاه معادلات

با جمع (*) و (**) :

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

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

$$\text{با قرار دادن } w_1 = \frac{1}{2} \text{ در (*)}$$

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

همچنین، چون در حدس اولیه $w_2 \propto (1, 0)$ داشتیم، باید صفر باشد تا ابرصفحه به مختصه x_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

(Support Vectors-SVs) بردارهای پشتیبان

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

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

معادلات ابرصفحه ± 1

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

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

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.

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

فاصله تا ابرصفحه تصمیم(حاشیه هندسی):

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

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 $2 = -2x_1$ و $x_1 = 4$ واحد است.

پرسش ششم)

الف. پیشنهاد مدل مناسب با استفاده از شبکه‌های عصبی

هدف ما مدل‌سازی مرزهای تصمیم‌پیچیده 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) 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 X و Y).

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

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

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

توابع پیشنهادی:

: ReLU یا Sigmoid

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، معمولاً انتخاب ارجح در لایه‌های میانی است.

نتیجه: برای طبقه‌بندی این داده‌های غیرخطی، یک MLP با تابع فعال سازReLU در لایه‌های مخفی

مناسب‌ترین انتخاب است.

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

1. ضعف MLP با تابع فعال ساز سراسری(Sigmoid/tanh)

اگرچه MLP به خودی خود می‌تواند مرزهای غیرخطی را مدل کند، استفاده از توابعی مانند Sigmoid یا 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 مانند tanh و Sigmoid 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): تابع فعال ساز سراسری به طور

ذاتی تمایل به ایجاد مرزهای تصمیم صاف و سراسری دارند. این مسئله باعث می‌شود مدل برای جدا کردن چهار

ناحیه خوش‌های کاملاً 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 مانند ReLU 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 (مانند SVM خطی یا MLP بدون لایه‌های مخفی غیر خطی) عملکرد بسیار ضعیفی در تفکیک این داده‌ها دارند زیرا:

فرض تفکیک پذیری 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 تعمیمیم بافته از مسئله XOR (Exclusive OR) در فضای بعد پایین است. برای طبقه‌بندی یک نقطه جدید در مرکز، خط جداکننده مجبور است یک کلاس (مثلاً آبی) 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. مدل‌های خطی این توانایی را ندارند.

پرسش پیاده سازی (RBF)

هدف این پرسشن، مدل سازی یک سیستم دینامیکی ساده با استفاده از یک شبکه عصبی باتابع شعاعی پایه است. در گام اول با پیاده سازی یک RBFNN استاندارد با ساختاری ایستا شروع می شود (بخش اول) و سپس یک شبکه پیشرفته تر و تطبیقی (بخش دوم) توسعه داده می شود که ساختار خود را بر اساس شبکه M-RAN که در [مقاله](#) ارائه شده، بهینه سازی می شود.

شناسایی سیستم با شبکه عصبی RBF

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

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

خلاصه پاسخ‌های تشریحی:

1.1.3 ماهیت مسئله: مسئله تقریب تابع (Function Approximation) است. شبکه تلاش می‌کند نگاشت بین ورودی‌های گذشته ($y(t-1), y(t-2), u(t-1)$) و خروجی فعلی ($y(t)$) را یاد بگیرد.

1.2.1 نقش μ_K و σ_K (مرکز): مختصاتی که نورون در انجا حداکثر پاسخ را دارد. σ_K (عرض): شعاع تاثیرگذاری محلی نورون را کنترل می‌کند. σ -کوچک \leftarrow پاسخ محلی و تیز. σ -بزرگ \leftarrow پاسخ گسترده و هموار.

1.2.2 تفسیر خروجی تست: اگر $\mu_K = x$ باشد، خروجی فعال‌ساز گوسی برابر با 1 است. این نشان‌دهنده حداکثر حساسیت و مشارکت نورون است.

1.3.4 محاسبه LLS و BP: RBFNN ایستا یک مسئله خطی در پارامترها است. Backpropagation نیاز به تکرارهای (Epochs) زیاد دارد. LLS راه حل بهینه kوچک است.

1.3.6 تاثیر K (تعداد مراکز): kوچک: منجر به کم‌برازش (Underfitting) می‌شود (ظرفیت مدل کافی نیست) و RMSE بالا می‌ماند. kبزرگ: به بیش‌برازش (Overfitting) می‌شود (حفظ نویز داده‌های آموزش) و RMSE_{Test} افزایش می‌یابد.

1.3.8 تاثیر σ (عرض): منجر به بیش برازش می‌شود (پوشش بسیار مجلی و تیز)، تعمیم‌پذیری کم است.

σ (بزرگ): منجر به کم برازش می‌شود (پوشش بسیار هموار)، مدل نمی‌تواند جزئیات را یاد بگیرد.

1.3.9 چالش‌های ساختار ایستا: چالش اصلی، وابستگی متقابل K و σ و انتخاب تصادفی مراکز است.

انتخاب تصادفی مراکز تضمین نمی‌کند که نواحی مهم فضای ورودی پوشش داده شوند.

1.4.2 نقش λ (Regularization): جریمه‌ای است که به نرم وزن‌های a اعمال می‌شود. اگر کوچک

باشد نزدیک به LLS بدون جریمه، خطر بیش برازش را افزایش می‌دهد. اگر بزرگ باشد وزن‌ها را به سمت

صفر سوق می‌دهد، منجر به کم برازش (مدل بیش از حد ساده) می‌شود.

1.5.3 مزیت K-Means: تضمین می‌کند که مراکز RBFNN در نواحی چگال (Dense Regions) فضای

ورودی قرار گیرند و به صورت متعادل و نمایانگر توزیع داده‌ها انتخاب شوند. این کار پوشش بهتری نسبت

به انتخاب کاملا تصادفی فراهم می‌کند.

2.1.2 ضرورت دو معیار رشد: معیار تازگی و خطأ هر دو لازمند. خطأ : منجر به رشد بیش از حد و

همپوشانی می‌شود (نورون‌های اضافی). تازگی: نورون‌هایی که ضرورتی ندارند اضافه می‌شوند (چون خطأ کم

بوده)، که پیچیدگی را افزایش می‌دهد.

2.2.5 مزایای M-RAN: مزیت: پیدا کردن توپولوژی مینیمال با دقت بالا (به دلیل رشد هوشمند و هوس).

توپولوژی مینیمال در کاربردهای دنیای واقعی (مانند سیستم‌های بلادرنگ و دستگاه‌های IoT) به دلیل

نیاز کمتر به حافظه و زمان استنتاج حیاتی است.

پرسش پیاده سازی (SVM)

گزارش حاشیه: عرض حاشیه نشان‌دهنده قائله بین ابرصفحه‌هایی است که کلاس‌ها را جدا می‌کنند (Hyperplanes). مدل‌های SVM خطی به دنبال حاشیه ماکزیمم هستند تا تعیین‌پذیری (Generalization) بهبود یابد.

تحلیل بایاس-واریانس: واریانس پائین و بایاس بال (Underfitting). حاشیه گسترده‌تر می‌شود و اجازه خطاهای اموزشی بیشتری را می‌دهد. واریانس بالا و بایاس پائین (Overfitting). حاشیه تنگ‌تر می‌شود و مدل تلاش می‌کند تمام نقاط را به درستی طبقه‌بندی کند، که منجر به افزایش W می‌شود.

تفسیر منحنی ROC: توانایی مدل در تفکیک کلاس‌ها (تشخیص صحیح و غلط) را در تمام استانه‌های تصمیم‌گیری نشان می‌دهد. AUC نزدیک به 1 نشان‌دهنده قدرت تفکیک عالی مدل است.

تفسیر منحنی Precision-Recall: منحنی برای ارزیابی مدل در مساوی با کلاس‌های نامتوزن (مانند تشخیص بیماری) بسیار مهم است. منحنی نزدیک به بالا-راست نشان‌دهنده دقیقت و حساسیت بالا به طور همزمان است، که در تشخیص پزشکی عالی است.

