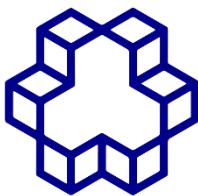


به نام خدا



دانشگاه صنعتی خواجه نصیرالدین طوسی

دانشکده برق

درس مبانی سیستم‌های هوشمند

مینی پروژه 2

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>

پرسش اول)

1. فرم اولیه SVM سخت حاشیه (Primal Form)

هدف: یافتن یک ابر صحنه جداکننده که کلاس‌ها را بطور کامل جدا کند 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 حاشیه 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:

$$(w \cdot x_+ + b = +1) \text{ برای بردارهای پشتیبان با } y_i = +1$$

$$(w \cdot x_+ + b = -1) \text{ برای بردارهای پشتیبان با } y_i = -1$$

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 چایگزینی $w \cdot (x_+ - x_-) = 2$ ، به نتیجه می رسیم:

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

3. استخراج فرم دوگانه و توجیه بردارهای پشتیبان

الف. استخراج فرم دوگانه (Dual Form)

برای استخراج فرم دوگانه، از ضرب کننده های لاگرانژ ($a_i \geq 0$) استفاده می کنیم.

1. تابع لاگرانژ (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 (Saddle Point)، \mathcal{L} را نسبت به w و b مشتق می گیریم

و بابر 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 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:

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

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) در تابع تصمیم ظاهر می‌شوند، شرط مکمل بودن (Slackness Complementary) از شرایط KKT 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 آنجایی که در فرم نهایی تابع تصمیم، $f(x_{new})$ یک مجموع وزن‌دار است که با a_i ضرب شده است،

فقط نمونه‌هایی 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).

$0 < \xi_i < 1$: 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 آن 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 آن $f(x_i) = w \cdot x_i + b$.

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 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:

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

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

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

کران پایین قرار دهیم:

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

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

4. جایگزینی: 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 $\max(0, 1 - y_i f(x_i))$ 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 تعمیم پذیری (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 طبقه بندی شوند (خطای آموزش افزایش یابد).

این مدل ساده تر است (نرم w کوچکتر است) و احتمال بیش برآزش (Overfitting) کاهش می‌یابد.

بنابراین، تعمیم پذیری 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 طبقه بندی کرده و نقض حاشیه را به حداقل برساند

(خطای آموزش کمینه می‌شود).

این مدل ممکن است پیچیده تر شود (نرم w بزرگتر است) و احتمال بیش برآزش افزایش می‌یابد.

تعمیم پذیری 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 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 ممکنه SVM سخت حاشیه نزدیک شود.

در نتیجه، تعداد بردارهای پشتیبان کاهش می‌یابد. (البته 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 تاکید بر دقت آموزش (مدل

پیچیده)، و C کوچک یعنی تاکید 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 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 $\xi_i \geq 0$ 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، از تابع لاگرانژ استفاده می‌کنیم:

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

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 b, w و ξ_i در \mathcal{L} ، فرم دوگانه به دست می‌آید:

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

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

ویژگی	ℓ_1 Soft Margin (جریمه خطی)	ℓ_1 Soft 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 به C محدود می شوند، که اغلب منجر به تعداد کمتری SV می شود.	a_i ها کران بالا ندارند، که اغلب منجر به تعداد بیشتری SV (و SV های با a_i کوچک) می شود.
وزن SV های خارج حاشیه	SV های خارج حاشیه ($\xi_i \leq 1$) همیشه $a_i = C$ دارند. (نقاط پرت به شدت جریمه و سخت محدود می شوند)	SV های خارج حاشیه ($\xi_i \geq 1$) می توانند $a_i > C$ داشته باشند. (نقاط پرت به صورت نرم تر و پیوسته تر جریمه می شوند.)

توضیح اثر بر w (وزن ها)

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

در هر دو حالت،

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

ℓ_2 : کران a_i ندارد. $\xi_i = \frac{a_i}{2C}$. این بدان معناست که a_i می تواند بزرگ باشد تا به یک نقطه پرت بزرگ پاسخ دهد. این جریمه پیوسته باعث می شود که $SVM - \ell_2$ نسبت به نقاط پرت حساس تر باشد و ممکن است w بزرگتر و متراکم تر 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 می شود، یک تابع ریاضی است که شرایط خاصی را ارضا می کند 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 می کند: $\phi: \mathcal{X} \rightarrow \mathcal{H}$.

شرط اصلی معتبر بودن (قضیه مرسر-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 می دهد بدون نیاز به تعریف و محاسبه صریح $\phi(x)$ (که اغلب

بسیار پیچیده یا بی نهایت بعد است)، ضرب داخلی $\phi(x_i) \cdot \phi(x_j)$ را در فضای بعد بالا محاسبه کنیم. این

تکنیک به ترفند کرنل (*Kernel Trick*) معروف است.

ب. خواص کرنل های *RBF* و چند جمله ای

کرنل های *RBF* و چند جمله ای دو مثال رایج از کرنل معتبر هستند:

کرنل چند جمله ای (Polynomial Kernel):

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

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 می دهد.

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 آن 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-Tucker-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 فعال 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)	$0 \leq a_i \leq C$	دوگان پذیری
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/شل شدگی

نقش هر حالت:

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

نقاط غیرپشتیبان (Non-SV):

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

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):

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

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

بردارهای پشتیبان خطا (SV violating Margin):

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

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

پرسش پنجم)

داده ها

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

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

الف. حدس ابر صفحه حداکثر حاشیه و معادله مرز تصمیم

حدس ابر صفحه

با مشاهده نقاط داده، متوجه می شویم که داده ها حول محور y 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 محور y ، حداقل 2 واحد است.

بنابراین، بهترین و متقارن ترین ابر صفحه جداکننده (خط جداکننده در \mathbb{R}^2) باید محور y باشد.

معادله مرز تصمیم (ابرصفحه)

معادله محور y به صورت زیر است:

$$x = 0 \text{ یا } 1.x + 0.y = 0$$

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:

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

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

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:

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

شناسایی بردارهای پشتیبان

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

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

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

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 (**):

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

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

با قرار 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_2 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

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 $w^T x + b = \pm 1$

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

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

ابرفحه حاشیه مثبت (+1):

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

ابرفحه 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):

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

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 x و y).

لایه مخفی: حداقل دو لایه مخفی با تعداد گره‌های کافی (مثلاً 10 گره در هر لایه) برای یادگیری ویژگی‌های غیر

خطی.

توابع فعال‌ساز (Activation Functions) لایه‌های مخفی: برای معرفی خاصیت غیرخطی، باید از توابع

فعال‌ساز غیرخطی مانند ReLU (Rectified Linear Unit) یا Sigmoid/tanh استفاده شود.

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

توابع پیشنهادی:

Sigmoid یا ReLU :

Sigmoid: در گذشته رایج بود، اما به دلیل مشکل شیب محوشونده (Vanishing Gradient) در لایه های

عمیق کمتر استفاده می شود.

ReLU: به دلیل سادگی و جلوگیری از مشکل شیب محوشونده، معمولاً انتخاب ارجح در لایه های میانی است.

نتیجه: برای طبقه بندی این داده های غیرخطی، یک MLP با توابع فعال ساز ReLU در لایه های مخفی

مناسب ترین انتخاب است.

ب. ضعف سایر مدل های شبکه عصبی و عملکرد ضعیف مدل های خطی

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

اگرچه MLP به خودی خود می تواند مرزهای غیرخطی را مدل کند، استفاده از توابعی مانند Sigmoid یا tanh

در لایه های مخفی ممکن است در این مسئله خاص (داده های خوشه ای) به طور موثر عمل نکند.

توابع فعال ساز سراسری (Global Activation Functions): توابعی مانند Sigmoid و tanh شیب های خود

را در نواحی دور از صفر از دست می دهند (شیب محوشونده). در یک معماری عمیق، این امر باعث می شود

یادگیری ویژگی ها کند شود.

مشکل کرزهای تصمیم غیر محلی (Non-local Decision Boundaries): توابع فعال ساز سراسری به طور

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

ناحیه خوشه ای کاملاً مجزا (مانند شکل 5 در سوالات) دچار مشکل شود و ممکن است مرزهای تصمیم، بیش از

حد از نقاط آموزش 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 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 تعمیم‌یافته از مسئله XOR (Exclusive OR) در فضای بعد پایین است. برای طبقه‌بندی یک نقطه جدید در مرکز، خط جداکننده مجبور است یک کلاس (مثلاً 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 آنجا به صورت خطی قابل تفکیک باشند. مدل‌های خطی این توانایی را ندارند.

پرسش پیاده سازی RBF

هدف این پرسش، مدل سازی یک سیستم دینامیکی ساده با استفاده از یک شبکه عصبی با تابع 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 توپ 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 : μ_K (مرکز): مختصاتی که نورون در آنجا حداکثر پاسخ را دارد. σ_K (عرض): شعاع

تأثیرگذاری 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 $x = \mu_K$ باشد، خروجی فعال‌ساز گوسی برابر با 1 است. این نشان‌دهنده حداکثر حساسیت و مشارکت نورون است.

1.3.4 محاسبه LLS و BP: LLS بسیار سریع است. زیرا RBFNN ایستا یک مسئله خطی در پارامترها است و LLS راه‌حل بهینه Backpropagation نیاز به تکرارهای (Epochs) زیاد دارد.

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

1.3.8 تاثیر σ (عرض): منجر به بیش برآزش می شود (پوشش بسیار مجلی و تیز)، تعمیم پذیری کم است.

σ (بزرگ): منجر به کم برآزش می شود (پوشش بسیار هموار)، مدل نمی تواند جزئیات را یاد بگیرد.

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

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

1.4.2 نقش λ (Regularization): جریمه ای است که به نرم وزن های a اعمال می شود. اگر کوچک

باشد نزدیک به LLS بدون جریمه، خطر بیش برآزش را افزایش می دهد. اگر بزرگ باشد وزن ها را به سمت

صفر سوق می دهد، منجر به کم برآزش (مدل بیش از حد ساده) می شود.

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

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

به انتخاب کاملاً تصادفی فراهم می کند.

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

همپوشانی می شود (نورون های اضافی). تازگی: نورون هایی که ضرورتی ندارند اضافه می شوند (چون خطا کم

بوده)، که پیچیدگی را افزایش می دهد.

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 سیستم های بلادرنگ و دستگاه های IOT) به دلیل

نیاز کمتر به حافظه و زمان استنتاج حیاتی است.

پرسش پیاده سازی SVM

گزارش حاشیه: عرض حاشیه نشان دهنده فاصله بین ابرصفحه‌هایی است که کلاس‌ها را جدا

می‌کنند (Hyperplanes). مدل‌های SVM خطی به دنبال حاشیه ماکزیمم هستند تا تعمیم‌پذیری

(Generalization) بهبود یابد.

تحلیل بایاس- واریانس: واریانس پائین و بایاس بال (Underfitting). حاشیه گسترده‌تر می‌شود و اجازه

خطاهای آموزشی بیشتری را می‌دهد. واریانس بالا و بایاس پایین (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 نشان دهنده قدرت تفکیک عالی مدل است.

تفسیر منحنی Precision-Recall: منحنی برای ارزیابی مدل در مسائل با کلاس‌های نامتوازن (مانند

تشخیص بیماری) بسیار مهم است. منحنی نزدیک به بالا- راست نشان دهنده دقت و حساسیت بالا به طور

همزمان است، که در تشخیص پزشکی عالی است.

