

University of Tehran

دانشکده‌گان علوم و فناوری های میان رشته ای



---

# Machine Learning

---

HW2

Fall 2024

## توضیحات مهم

سلام بر دانشجویان عزیز، چند نکته مهم:

۱. حجم گزارش به هیچ عنوان معیار نمره‌دهی نیست، در حد نیاز توضیح دهید.
۲. نکته‌ی مهم در گزارش‌نویسی روشن بودن پاسخ‌ها می‌باشد، اگر فرضی برای حل سوال استفاده می‌کنید حتماً آن را ذکر کنید، اگر جواب نهایی عددی است به‌صورت واضح آن را بیان کنید.
۳. برای سوالات شبیه‌سازی، فقط از دیتاست داده‌شده استفاده کنید.
۴. فایل نهایی خود را در یک فایل زیپ شامل گزارش با فرمت PDF، آپلود کنید. نام فایل زیپ ارسالی باید الگوی زیر را داشته باشد :

ML\_HW1\_StudentNumber

۵. کد سوالات شبیه‌سازی بصورت فایلی تحت عنوان سوال آن با پسوند py یا ipynb به همراه گزارش در فایل زیپ تحت الگوی بند ۴ پیوست شود.
۶. هرگونه شباهت در گزارش و کد مربوط به شبیه‌سازی، به منزله‌ی تقلب می‌باشد و کل تمرین برای طرفین صرف‌نظر خواهد شد.
۷. تمامی سوالات خود را به دستیار آموزشی تمرین مربوطه به ایمیل زیر ارسال نمایید:

zeinab.yazdani@ut.ac.ir

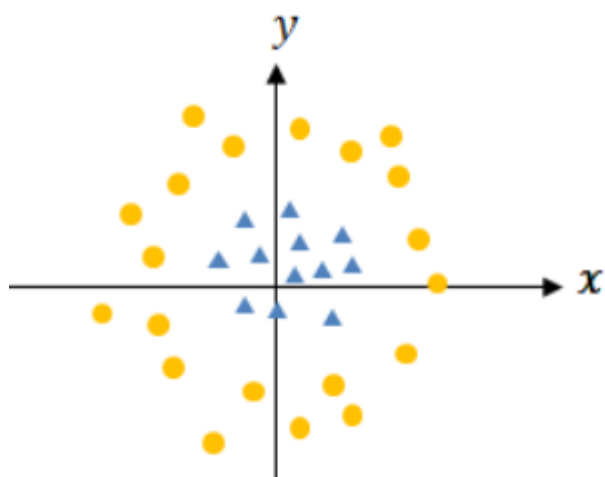
## سوال اول: SVM

**الف.** تفاوت hard margin و soft margin در SVM چیست؟ مثال ساده را در فضای دو بعدی نشان دهید که استفاده از soft margin به جای hard margin در ازای وجود تعدادی خطا، منجر به بهبود طبقه‌بند خواهد شد.

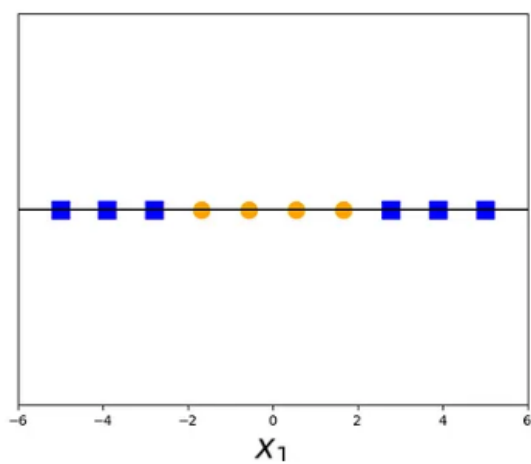
**ب.** کاربرد و مفهوم ضریب  $c$  در معادله مربوط با soft margin چیست و افزایش و کاهش آن چه تاثیری در طبقه‌بند خواهد داشت؟

$$\max \frac{1}{2} \|w\|^2 + c \sum_{i=1}^n \xi_i$$

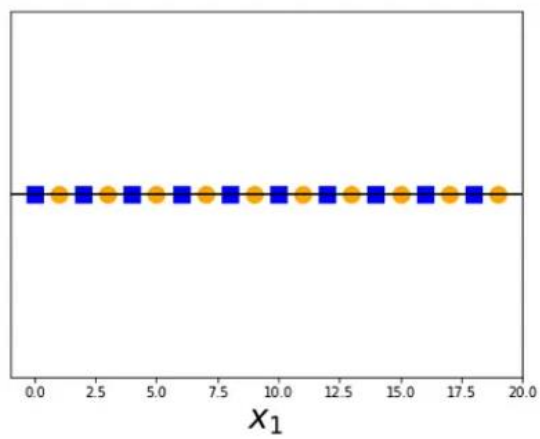
**ج.** مفهوم کلی کرنل و دلایل استفاده از روش‌های مبتنی بر کرنل را بیان کنید و برای هر یک از مجموعه-داده‌های زیر یک کرنل پیشنهاد کنید.



شکل ۱- سوال ۳، قسمت ج، مجموعه داده ۱



شکل ۲- سوال ۳، قسمت ج، مجموعه داده ۲



شکل ۳- سوال ۳، قسمت ج، مجموعه داده ۳

## سوال دوم: MLP

فرض کنید یک شبکه  $MLP^1$  سه لایه شامل یک لایه ورودی، یک لایه پنهان با ۳ نورون یک لایه خروجی با یک نورون داریم. از تابع فعال‌ساز<sup>۲</sup> سیگموئید<sup>۳</sup> برای لایه پنهان و تابع همانی در لایه آخر استفاده کنید و تابع هزینه<sup>۴</sup> را  $MSE^5$  در نظر بگیرید.

ورودی:

$$x = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

وزن‌های لایه پنهان

$$w1 = \begin{bmatrix} 0 & 1 & 2 \\ 0 & -1 & 0 \end{bmatrix}$$

بایاس: ۱

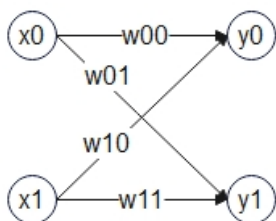
وزن‌های لایه آخر:

$$w2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

بایاس: ۰

برای راهنمایی در مورد نحوه نام‌گذاری‌ها از شکل زیر استفاده کنید.

$$\begin{bmatrix} x0 \\ x1 \end{bmatrix} \quad \begin{bmatrix} w00 & w01 \\ w10 & w11 \end{bmatrix}$$



---

<sup>1</sup> Multi Layer Perceptron

<sup>2</sup> Activation Function

<sup>3</sup> Sigmoid

<sup>4</sup> Loss Function

<sup>5</sup> Cross Entropy

**الف.** فرض کنید برچسب واقعی این داده ۱ باشد، یک مرحله از به‌روزرسانی وزن‌ها را از طریق الگوریتم پس‌انتشار خطا<sup>۱</sup> به صورت کامل و مرحله به مرحله با محاسبات مربوطه انجام داده و گرادیان نسبت به تمام پارامترها را به دست بیاورید. نرخ یادگیری را ۰.۵ در نظر بگیرید.

(حل سوال با استفاده از روابط ماتریسی علاوه بر افزایش سرعت و دقت و تعمیم‌توانایی شما به حل مسائل با تعداد نوروں و لایه بیشتر و ورودی‌های بزرگتر، **نمره امتیازی** نیز خواهد داشت).

**ب.** دلیل استفاده از توابع فعال‌ساز در شبکه‌های عصبی چند لایه چیست؟

**ج.** دو مورد از مشکلات سیگموئید را توضیح دهید و برای آنها راه‌حل‌هایی ارائه کنید. به نظر شما دلیل استفاده از تابع فعال‌ساز ReLU به جای سیگموئید چیست؟

**د.** توضیح دهید در مسیر پیشرو و پسرو<sup>۲</sup> با تابع ReLU چگونه برخورد می‌شود (گرادیان خروجی ReLU نسبت به ورودی آن).

---

<sup>۱</sup> Backpropagation

<sup>۲</sup> Forward and Backward Path

## سوال سوم (پیاده‌سازی): SVM

به همراه فایل‌های پیوست در صورت تمرین، یک مجموعه داده پزشکی برای طبقه‌بندی بیماری دیابت ارائه شده است. مشخصات این داده‌ها در جدول ۱ بیان شده است. در این سوال با استفاده از SVM یک طبقه‌بندی برای تشخیص این بیماری طراحی خواهد شد.

جدول ۱- مشخصات مجموعه داده تشخیص دیابت

Pregnancies	Number of times pregnant
Glucose	Plasma glucose concentration a 2 hours in an oral glucose tolerance test
BloodPressure	Diastolic blood pressure (mm Hg)
SkinThickness	Triceps skin fold thickness (mm)
Insulin	2-Hour serum insulin (mu U/ml)
BMI	Body mass index (weight in kg/(height in m)^2)
DiabetesPedigreeFunction	Diabetes pedigree function
Age	Age (years)
Outcome	Class variable (0 or 1) 268 of 768 are 1, the others are 0

## الف. EDA (Exploratory Data Analysis)

در برخورد با مجموعه داده‌های مختلف، بسیار مهم است که شما بتوانید به صورت دیداری اطلاعاتی را از ظاهر مجموعه داده کسب کنید تا بتوانید برای مراحل بعدی برنامه ریزی مناسب تری داشته باشید. در این قسمت برای درک بهتر دادگان، سعی کنید آن‌ها را با ابزارهای مختلف نمایش دهید و به صورت ظاهری و نیز از نظر آماری ویژگی‌های مختلف و ارتباط آنها با خروجی را بررسی کنید. نمودار scatter plot هر دوتایی از ویژگی‌ها را ترسیم کنید. توجه داشته باشد که نمودار ترسیم شده باید با درج تمامی اطلاعات مورد نیاز (برچسب مناسب برای محور ها، عنوان مناسب برای هر نمودار و ...) همراه باشد. توجه کنید که تحلیل نمودار ها در این سوال اهمیت بالایی دارد. بنابراین در این قسمت پس از خواندن داده‌ها سعی کنید مشخصات آن را بررسی کنید (وجود داده‌های گم‌شده<sup>۱</sup>، بررسی پارامترهای آماری و تصویر سازی و رسم رابطه هر ویژگی با خروجی و...) و بررسی کنید کدام ویژگی‌ها برای تصمیم گیری مفید تر هستند..

<sup>1</sup> Missing data



## ب. پیش‌پردازش

یکی از مراحل مهم در برخورد با داده‌های دنیای واقعی، مرحله پیش‌پردازش است. در مورد پیش‌پردازش‌های معمول قبل از استفاده از داده‌های خام تحقیق کنید. با ذکر دلیل بیان کنید انجام چه پیش‌پردازش‌هایی روی داده‌های این سوال به مسئله کمک می‌کند و این پیش‌پردازش‌ها را اعمال کنید. انجام درست قسمت قبل، در این قسمت به شما کمک زیادی می‌کند.

## ج. بررسی راه‌حل (نیازی به پاسخ طولانی نیست، تنها رساندن مفهوم کافی‌ست)

۱. در مورد مفاهیم Grid Search و Random Search تحقیق کنید و هر کدام را مختصراً توضیح دهید.

۲. در مورد کرنل‌های مختلف مانند Linear, RBF, Polynomial تحقیق کنید و یک یا دو مورد از مهم‌ترین پارامترهای هر کدام را مختصر توضیح دهید.

۳. روش‌های one vs all و one vs rest را مختصراً توضیح داده و باهم مقایسه کنید. آیا در این مسئله نیازی به استفاده از آنها داریم؟

## د. طبقه‌بندی

۱. داده‌ها را به دو بخش آموزش و آزمون تقسیم کنید.

۲. مدل svm خود را ایجاد کرده و با استفاده از Grid Search از بین سه کرنل معرفی شده در قسمت قبل، بهترین پارامترهایی (c, gamma و d) که برای هر کرنل بدست می‌آید را روی داده‌های آموزش گزارش کرده و برای هر کدام score را بیان کنید.

## ه. نتایج

بهترین مدل را روی داده‌های آموزش و آزمون اعمال کنید و نتایج را گزارش کنید.

## سوال چهارم (پیاده سازی): MLP

در این سوال، با استفاده از کتابخانه **Scikit-learn** یک شبکه عصبی چند لایه برای طبقه‌بندی تصاویر miniMNIST که شامل تصاویر ۱۰۰۰۰ رقم دست‌نویس است، طراحی خواهید کرد. فایل مربوط به این مجموعه داده همراه با سوالات در اختیار شما قرار داده شده است. برای پیاده‌سازی باید از همین فایل استفاده کنید.

**الف.** مجموعه داده ارائه شده به صورت یک فایل `mat` است که شما باید کتابخانه مناسب برای خواندن آن را پیدا کرده و پس از خواندن تصاویر را از برچسب‌ها جدا کنید. `shape` تصاویر و برچسب‌ها را بررسی کنید و تغییرات مورد نیاز را اعمال کنید.

**ب.** برای آشنایی با مجموعه داده مورد نظر، یک مجموعه تصادفی ۵۰ تایی از داده‌ها را به صورت ۵ در ۱۰ نمایش دهید.

**ج.** داده‌ها را به دو بخش آموزش (۸۰٪) و آزمون (۲۰٪) تقسیم کنید. توضیح دهید دلیل تقسیم مجموعه داده هر کدام چه کاربردی دارند؟ عدم استفاده از مجموعه اعتبارسنجی در این سوال منجر به چه مشکلی خواهد شد؟

**د.** یک شبکه عصبی با یک لایه پنهان شامل ۱۶ نورون پیاده‌سازی کنید.

۱. تعداد نورون‌ها را به ۱۲۸ نورون تغییر دهید و نتایج را مقایسه کنید.

۲. تعداد لایه‌های پنهان را به دو لایه (اولی ۶۴ و دومی ۱۲۸ نورون) تغییر دهید و نتایج را مقایسه کنید.

**د.** در مورد بیش برازش<sup>۱</sup> و کم برازش<sup>۲</sup> در شبکه‌های عصبی توضیح دهید و برای پیشگیری و حل این مشکلات راهکارهایی پیشنهاد کنید (برای هر کدام دو مورد کافیست). نتایج قسمت‌های قبل را با توجه به این مفاهیم تحلیل کنید و بیان کنید در هر مرحله با کدام یک از این مشکلات مواجه بودید.

---

<sup>۱</sup> Overfit

<sup>۲</sup> Underfit

## سوال پنجم:

الف. در یک مسئله تشخیص الگو با چهار کلاس، توابع تشخیص زیر را داریم :

$$\text{class 1: } g_1(x, y) = x - y + 1$$

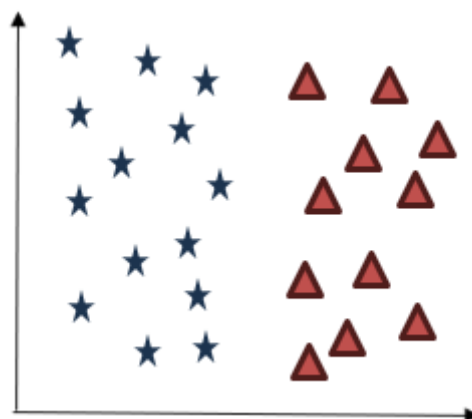
$$\text{class 2: } g_2(x, y) = -x - y + 2$$

$$\text{class 3: } g_3(x, y) = -x + y - 3$$

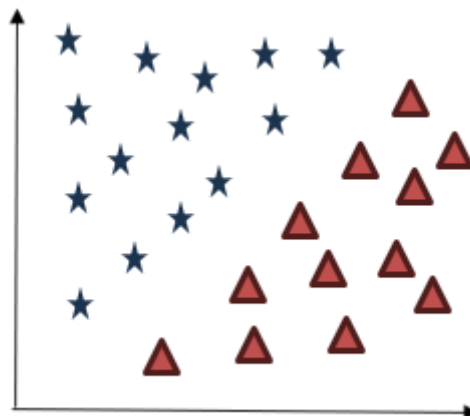
$$\text{class 4: } g_4(x, y) = x + y$$

مرز تصمیم‌گیری را پیدا کرده و ترسیم کنید، با این قاعده طبقه‌بندی که برای یافتن برچسب کلاس از حداکثر مقدار  $g_i$  برای هر نقطه داده ورودی  $(x, y)$  استفاده کند.

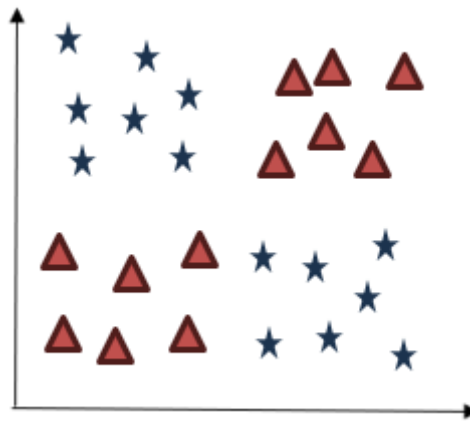
ب. نمودارهای زیر نشان‌دهنده نمونه‌های آموزشی برای یک تابع مفهوم با مقادیر بولی هستند که در فضای ویژگی رسم شده‌اند. در هر مورد، چه تعداد گره (neuron) برای طبقه‌بندی صحیح توسط یک پرسپترون مورد نیاز است؟



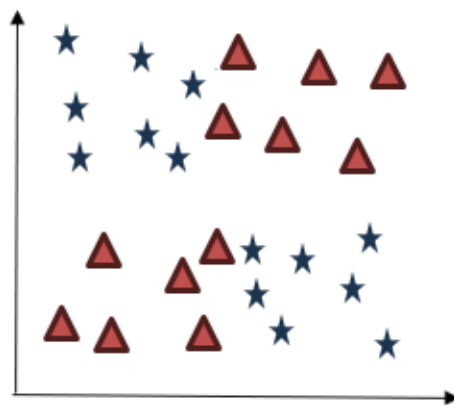
تعداد گره:



۱. تعداد گره:



۲. تعداد گره:

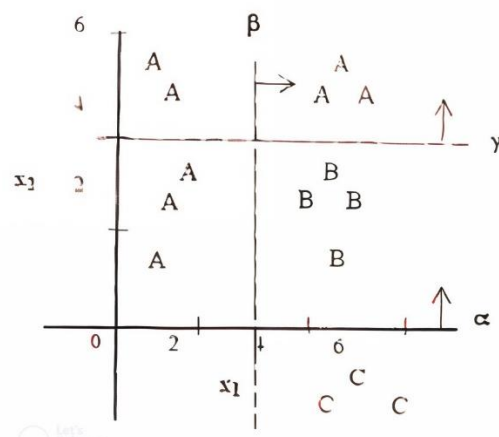


۳. تعداد گره:

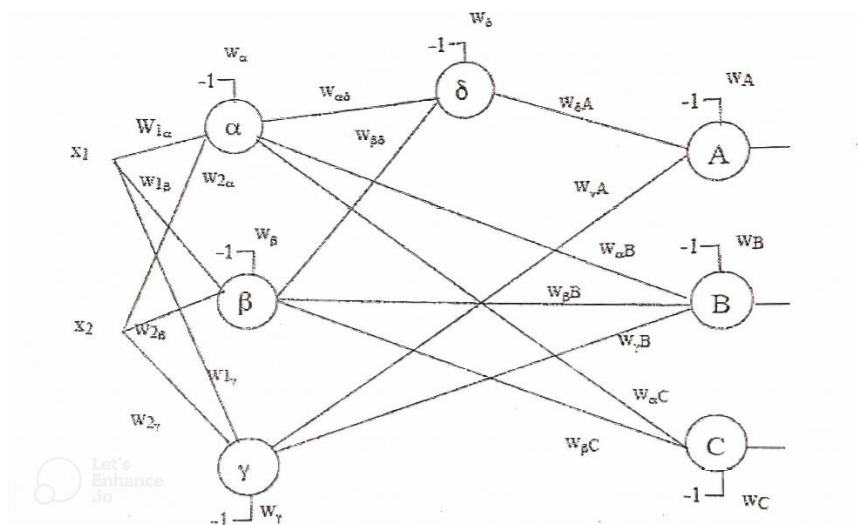
ج.

(۱) یک شبکه عصبی را در نظر بگیرید که ورودی‌های  $X_1, X_2, \dots, X_n$  دارد، که هر کدام می‌توانند ۰ یا ۱ باشند. این شبکه با تعیین وزن‌های روی پیوندها و تابع فعال‌سازی  $g$  در گره مشخص می‌شود. شبکه‌ای طراحی کنید که "تابع اکثریت" را برای  $n$  گره ورودی محاسبه کند. یک تابع اکثریت باید در صورتی که حداقل نیمی از ورودی‌ها بالا (۱) باشند، خروجی ۱ تولید کند، و در غیر این صورت خروجی ۰ تولید کند.

(۲) وزن‌های گم‌شده را برای هر یک از گره‌ها در شبکه پرسپترون زیر پر کنید. فرض‌های زیر را در نظر بگیرید:



- خروجی‌های پرسپترون یا ۰ یا ۱ هستند.
- A, B و C نشان‌دهنده کلاس‌ها هستند.
- خط‌های  $\alpha$ ,  $\beta$  و  $\gamma$  نشان‌دهنده مرزهای تصمیم‌گیری هستند.
- جهت فلش‌های نشان داده شده روی نمودارها، سمتی از هر مرز را نشان می‌دهد که باعث می‌شود پرسپترون خروجی ۱ تولید کند.



$a$

$w_{1a}$	
$w_{2a}$	1
$w_a$	

$\beta$

$w_{1\beta}$	1
$w_{2\beta}$	
$w_{\beta}$	

$\gamma$

$w_{1\gamma}$	
$w_{2\gamma}$	1
$w_{\gamma}$	

$\delta$

$w_{a\delta}$	1
$w_{\beta\delta}$	
$w_{\delta}$	

$A$

$w_{\delta A}$	1
$w_{\gamma A}$	
$w_A$	

$B$

$w_{aB}$	1
$w_{\beta B}$	1
$w_{\gamma B}$	
$w_B$	

$C$

$w_{aC}$	
$w_{\beta C}$	1
$w_C$	