سلام بر دانشجویان عزیز، چند نکته مهم:

۱. حجم گزارش به هیچ عنوان معیار نمره دهی نیست، در حد نیاز توضیح دهید.

۲. نکته‌ی مهم در گزارش نویسی روشن بودن پاسخ‌ها می‌باشد، اگر فرضی برای حل سوال استفاده می‌کنید حتما آن را ذکر کنید، اگر جواب نهایی عددی است به صورت واضح آن را بیان کنید.

۳. کد های نوشته شده برای هر سوال شبیه سازی را در فایل ipynb متناظر آن سوال بنویسید.

٤. کدهای ارسال شده **بدون گزارش و یا کامنت گذاری دقیق در کد** فاقد نمره می‌باشند.

٥. برای سوالات شبیه سازی، فقط از دیتاست داده شده استفاده کنید.

٦. نمره تمرین از ۱۰۰ نمره می‌باشد

۷. هرگونه شباهت در گزارش و کد مربوط به شبیه سازی، به منزله تقلب می‌باشد و کل تمرین برای طرفین **۱۰۰-** خواهد شد.

۸. در صورتی که تشخیص داده شود از چت بات ها به صورت مستقیم برای پاسخ سوال های تئوری و شبیه سازی استفاده شده است، نمره **۱۰۰-** در نظر گرفته خواهد شد.

۹. فایل نهایی خود را در یک فایل زیپ شامل، pdf گزارش و فایل کدها آپلود کنید. نام فایل زیپ ارسالی الگوی ML_HW#_StudentNumber داشته باشد.

در صورت داشتن سوال، از طریق گروه درس یا ایمیل‌های زیر با تدریسیار مربوطه سوال‌های خود را مطرح کنید.

سوال ۱ و ۲ و ۶ : fatemehra10@gmail.com ، سوال ۳ و ۴ و ۷ : Kermaninia@ut.ac.ir
سوال ۵ و ۸ : mahmoos2078@gmail.com

سوال ۱ (نمره ۱۰):

به سوالات زیر پاسخ دهید:

الف) بالا بودن بایاس و یا بالا بودن واریانس در یک مدل نشان دهنده چه چیزی می باشد؟ با استفاده از چه روش هایی می توان میزان بایاس و واریانس مدل را کنترل کرد؟

ب) فرض کنید با استفاده از یک مدل رگرسیون، مدلی را ایجاد کرده ایم و میزان خطای مدل بر روی داده های آموزش پایین می باشد در صورتی که خطای مدل بر روی دادگان تست بسیار زیاد می باشد، علت این موضوع چیست؟ چه تغییراتی را می توان اعمال نمود تا این مشکل تا حدودی برطرف گردد؟

پ) یک مدل رگرسیون خطی Ridge به صورت زیر در نظر بگیرید که در آن $\lambda > 0$ پارامتر regularization می باشد:

$$w_{min} = \arg min_w \|y - Xw\|_2^2 + \lambda\|w\|_2^2$$

ثابت کنید فرم بسته w_{min} به صورت زیر می باشد :

$$w_{min} = (X^T X + \lambda I)^{-1} X^T y$$

ت) فرض کنید داده ها از فرم خطی زیر پیروی می کنند:

$$y = Xw^* + \varepsilon$$

که w^* بردار وزن های واقعی می باشد و $\varepsilon \sim N(0, \sigma^2 I)$ یک نویز گاوسی می باشد. با استفاده از فرم بسته به دست آمده w_{min} در مرحله قبل ، ثابت کنید مقدار امید ریاضی و واریانس w_{min} به صورت زیر می باشد:

$$(A = (X^T X + \lambda I)^{-1} X^T)$$

$$E[w_{min}] = (X^T X + \lambda I)^{-1} X^T X\, w^*$$

$$Var[w_{min}] = A\sigma^2 A^T$$

سوال ۲ (نمره ۱۵):

موارد زیر را اثبات کرده و به طور کامل توضیح دهید:

الف) فرض کنید در یک مدل رگرسیون خطی، متغیر هدف y به صورت زیر مدل سازی شده است:

$$y_i = w^T x_i + \varepsilon_i \, , \varepsilon_i \sim N(0, \sigma^2)$$

با فرض گاوسی بودن نویز ها و مستقل بودن آن ها، وزن هایی که با استفاده از روش بیشینه سازی احتمال (*MLE*)[1] بدست می‌آید همان وزن هایی است که با استفاده از تابع هزینه *SSE*[2] بدست می آوریم.

ب) در مدل *logistic regression*، احتمال $y \in \{0, 1\}$ برای ورودی x توسط تابع زیر مدل می شود:

$$P(y|x) = \sigma(w^T x) = \frac{1}{1 + \exp(-w^T x)}$$

ثابت کنید وزن هایی که با استفاده از روش بیشینه سازی احتمال (*MLE*) بدست می‌آید همان وزن هایی است که با استفاده از تابع هزینه binary cross entropy بدست می آوریم.

* در هر دو قسمت الف و ب فرض کنید برای بدست آوردن وزن ها بهینه سراسری آن را بدست می آوریم.

[1] Maximum likelihood
[2] Sum squared error

سوال ۳ (نمره ۱۵):

* تمام محاسبات این سوال باید بطور دستی انجام شود و نیازی به پیاده سازی با کد نیست.

مجموعه داده‌ای در فضای $\mathbb{R}^1$ با یک ویژگی X_1 داریم که کلاس‌های متناظر با آن‌ها -1 و $+1$ است. این مجموعه داده شامل سه نمونه‌ی $X_1 = \{-3, -2, 3\}$ از کلاس $+1$ و سه نمونه‌ی $X_1 = \{-1, 0, 1\}$ از کلاس -1 است.

الف) یک feature map ساده بصورت $\varphi(u) = (u, u^2)$ تعریف می‌کنیم که نقاط را از فضای $\mathbb{R}^1$ به $\mathbb{R}^2$ می برد. این feature map را روی داده‌ها اعمال کنید و نقاط را در فضای ویژگی اولیه و جدید رسم کنید. آیا این مجموعه داده در فضای ویژگی اولیه می‌توانست با یک جدا کننده‌ی خطی به طور کامل تفکیک شود؟ در فضای ویژگی جدید چطور؟

ب) فرم تحلیلی تابع کرنل $k(x_1, x_1')$ که متناظر با feature map اعمال شده (φ) است را بنویسید.

ج) خطی پیدا کنید که margin تشکیل شده بین آن و دو کلاس مختلف بیشینه باشد. معادله ی نرمال این خط را که به فرم $w_1 Y_1 + w_2 Y_2 + c = 0$ است، بنویسید (Y_1 و Y_2 ابعاد فضای ویژگی جدید و حاصل اعمال کردن feature map روی فضای ویژگی اولیه بصورت $\varphi(X_1) = (Y_1, Y_2)$ هستند.) و مقدار margin را حساب کنید.

* برای یافتن c و w_2 و w_1، با توجه به محل قرارگیری نقاط، از شهود هندسی و مفاهیم SVM استفاده کنید. (لازم به حل معادله‌ی quadratic نیست.)

د) خط یافت شده بعنوان decision boundary را به همراه Plus plane و Minus plane رسم کرده و Support vectorها را نشان دهید و بررسی کنید که تمام نقاط مجموعه‌ی داده‌ها به درستی کلاس بندی می‌شوند. سپس decision boundary حاصل از این خط را در همان فضای ویژگی اولیه ($\mathbb{R}^1$) نیز رسم کرده و درستی آن را بررسی کنید.

* راهنمایی: اگر تعداد support vectorهایی که یافته‌اید ۲ تا نشد، در پاسخ این بخش و بخش قبلی تجدید نظر کنید.

ه) می دانیم که فرم کلی مسئله‌ی دوگان برای hard SVM با اعمال کرنل، بصورت زیر است:

$$\max.\ W(\boldsymbol{\alpha}) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1,j=1}^{n} \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j)$$

Subject to:

$$\alpha_i \geq 0 \quad \text{and} \quad \sum_{i=1}^{n} \alpha_i y_i = 0$$

روند کلی بدست آمدن این عبارت را بدون وارد شدن به جزییات ریاضی بنویسید و بگویید هر متغیر نشانگر چیست و چرا Support vectorها نقش اساسی‌ای در پیدا کردن Decision boundary دارند.

و) حالا به کمک مختصات Support vectorها و تابع کرنلی که در بخش های قبلی یافتید و با اعمال قیدهای مسئله‌ی دوگان، تمام ضرایب مجهول عبارت بخش ه (α_i ها) را بدست آورید.

ز) می دانیم برای کلاس بندی نمونه‌های جدید، می‌توان از رابطه‌ی زیر استفاده کرد. نحوه‌ی بدست آمدن آن را توضیح دهید و با جایگذاری ضرایب محاسبه شده، b را بدست آورید. سپس نشان دهید مرز تصمیمی که با جایگذاری ضرایب بدست آمده در بخش قبلی و همین بخش در این رابطه بدست می‌آید، همان است که بصورت هندسی در بخش د بدست آوردید.

$$y(x) = \text{sign}\left(\sum_{n=1}^{|SV|} \alpha_n y_n k(x, u_n) + b\right)$$

سوال ۴ (نمره ۱۰):

❋ تمام محاسبات این سوال باید بطور دستی انجام شود و نیازی به پیاده سازی با کد نیست.

الف) می دانیم که مسئله ی دوگان Soft SVM به شکل زیر است. تفاوت آن را با مسئله‌ی دوگان Hard SVM و دلیل این تفاوت را بطور خلاصه توضیح دهید.

$$\max.\ W(\boldsymbol{\alpha}) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1,j=1}^{n} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j$$
$$\text{subject to } C \geq \alpha_i \geq 0, \sum_{i=1}^{n} \alpha_i y_i = 0$$

ب) با توجه به شکل زیر، ضریب α_i متناظر با هرکدام از نقاط زیر کدام حالت از حالات زیر

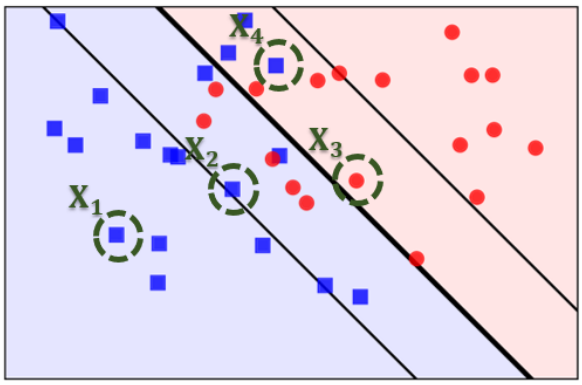
۱- $0 < \alpha_i < C$

۲- $\alpha_i = C$

۳- $\alpha_i = 0$

را خواهد داشت.

- نقطه‌ای مانند X_1 که در سمت درستِ مارجین قرار دارد.
- نقطه‌ای مانند X_2 که روی خود مارجین قرار دارد.
- نقطه‌ای مانند X_3 که در سمت اشتباه مارجین قرار دارد.
- نقطه‌ای مانند X_4 که در سمت اشتباه مرز تصمیم قرار دارد.



ج) حالا فرض کنید که نمی‌توانیم نقاطی مثل X_4 که کلاس‌بندی آن‌ها خطا دارد را داشته باشیم و می‌خواهیم کاری کنیم که بتوانیم در هر مجموعه‌ی متناهی از نقاط، تمام داده‌هایمان را بطور خطی جداسازی کنیم. پس از یک feature map به شکل زیر استفاده می‌کنیم که نقاط را از فضای $\mathbb{R}^1$ به فضای $\mathbb{R}^\infty$ می‌برد و ما را به هدفمان می‌رساند (با ساختن بردارهایی که مستقل خطی هستند). اما آیا می‌توان با ابزارهای محاسباتی موجود، این mapping را مستقیما روی داده‌ها اعمال کرده و بطور صریح $\phi_\infty(x)$ را بسازیم؟

$$\phi_\infty(x) = \left\{ e^{-x^2/2}, e^{-x^2/2}x, \frac{e^{-x^2/2}x^2}{\sqrt{2}}, \ldots, \frac{e^{-x^2/2}x^i}{\sqrt{i!}} \ldots \right\}$$

د) فرم بسته‌ی تابع کرنل متناظر با نگاشت تعریف شده در بالا را بدست آورید.

* راهنمایی: از بسط تیلور e^x استفاده کنید

$$\mathrm{e}^x = \lim_{n\to\infty} \sum_{i=0}^{n} \frac{\mathrm{x}^\mathrm{i}}{i!}$$

سوال ۵ (نمره ۱۰):

به سوالات زیر در مورد درخت تصمیم پاسخ دهید.

الف) درخت تصمیم را به روش ID3 با نوشتن محاسبات طبق داده های جدول ۱ آموزش دهید و آن را رسم کنید. دقت درخت تصمیم آموزش دیده به دست آمده را روی دادگان آموزش (جدول ۱) و آزمون (جدول ۲) به دست آورید.(ستون هدف بازی کردن یا نکردن است.)

	Outlook	Temperature	Wind	Play Tennis
1	Sunny	Hot	Weak	No
2	Sunny	Hot	Strong	No
3	Overcast	Mild	Weak	Yes
4	Rain	Mild	Weak	Yes
5	Rain	Cool	Weak	Yes
6	Rain	Cool	Strong	No
7	Overcast	Hot	Strong	Yes

جدول ۱ : دادگان آموزش

	Outlook	Temperature	Wind	Play Tennis
1	Sunny	Mild	Strong	No
2	Rain	Cool	Weak	Yes
3	Overcast	Hot	Weak	Yes
4	Rain	Cool	Strong	Yes

جدول ۲: دادگان آزمون

ب) روش آموزش درخت تصمیم C4.5 را بیان کنید و دو مزیت این روش نسب به روش ID3 بیان کنید.

سوال ۶ (شبیه سازی، نمره ۱۵):

در این تمرین، قصد داریم با استفاده از الگوریتم **لجستیک رگرسیون با *Regularization* نوع *L2***، مدلی برای تشخیص رانندگان حرفه‌ای از روی داده‌های شبیه‌سازی‌شده (Q6-drivers) طراحی کنیم. این داده‌ها شامل اطلاعات مربوط به عملکرد رانندگان در یک محیط شبیه‌سازی‌شده می باشند: (برای تمامی بخش ها علاوه بر توضیحات کلی کد نیاز است تحلیل خود را از نتایج به دست آمده ارائه دهید.)

- *reaction_time*: زمان واکنش (برحسب ثانیه)
- *steering_deviation*: میزان انحراف از مسیر مستقیم (درجه)
- *pro_driver*: آیا راننده حرفه‌ای است یا خیر (۱ برای حرفه‌ای، ۰ برای غیرحرفه‌ای)

بخش اول: پیش‌پردازش داده‌ها

الف) تعداد و درصد مقادیر گمشده در هر ویژگی را محاسبه کرده و نمایش دهید.

ب) برای مقادیر گمشده، یک روش مناسب (میانگین، میانه یا حذف) انتخاب و بر روی مقادیر گمشده اعمال نمایید و دلیل انتخاب روش خود را توضیح دهید.

پ) ویژگی‌ها را با یکی از روش‌های *standard scaling* یا *min-max scaling* نرمال‌سازی کنید.

ت) توزیع هر ویژگی را قبل و بعد از نرمال‌سازی با استفاده از هیستوگرام رسم و تحلیل نمایید.

بخش دوم: تحلیل اولیه داده‌ها

الف) نمودار *scatter* بین *reaction_time* و *steering_deviation* را رسم کرده و رنگ‌گذاری را بر اساس *pro_driver* انجام دهید.

ب) بررسی کنید که آیا داده‌ها به صورت خطی قابل جداسازی هستند؟ آیا می توان با استفاده از مدل *logistic regression* داده های دو کلاس را از یکدگیر جدا کرد؟ تحلیل خود را ارائه دهید.

بخش سوم: پیاده سازی مدل

برای بهبود جداپذیری، ویژگی‌ها را به فضای مرتبه بالاتر با استفاده از تابع زیر منتقل کنید:

$$f(x_1, x_2) = [x_1, x_2, {x_1}^2, {x_2}^2, x_1 x_2, {x_1}^3, {x_2}^3, {x_1}^2 x_2, x_1 {x_2}^2, \dots, {x_1}^5, {x_2}^5]$$

الف) داده‌ها را به نسبت ٪۷۰ آموزش و ٪۳۰ آزمون تقسیم کنید.

ب) یک مدل *Logistic Regression* با *Regularization* نوع *L2* پیاده‌سازی کنید.

ج) مدل را با استفاده از **گرادیان نزولی** آموزش داده و تغییرات تابع هزینه را رسم کنید.

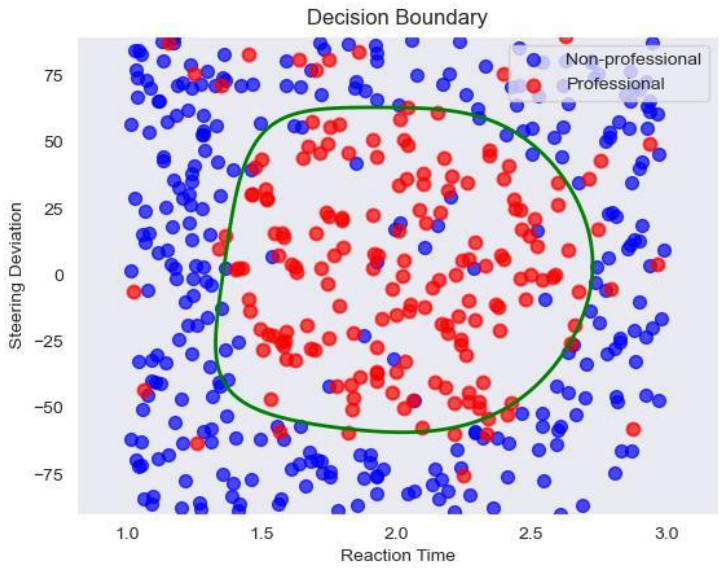
بخش چهارم: ارزیابی مدل

الف) با استفاده از داده‌ی آزمون، عملکرد مدل را با معیارهای زیر بسنجید.

- *Confusion Matrix*
- *Accuracy*
- *Precision*
- *Recall*
- *F1-score*

ب) نمودار *ROC* را رسم کرده و مقدار *AUC* را محاسبه و تحلیل نمایید.

ج) مرز تصمیم‌گیری *(Decision Boundary)* به دست آمده توسط مدل را رسم کنید. (خروجی شبیه به شکل زیر می باشد)



بخش پنجم: تحلیل حساسیت نسبت به مقدار λ (*Regularization*)

الف) مدل *Logistic Regression* را با مقادیر مختلف λ (۰ ، ۵۰ ، ۱۰۰، ۵۰۰) آموزش دهید.

ب) عملکرد مدل را با معیارهایی مانند *Accuracy*، *F1-score* و *AUC* برای هر λ محاسبه نمایید و نتایج را بر روی نمودار نمایش دهید.

ج) مرز تصمیم گیری را به ازای تمایی مقادیر رسم کرده و تحلیل کنید که چگونه افزایش یا کاهش λ روی مدل تأثیر می‌گذارد.

* در این سوال پیاده سازی مدل logistic regression و هم چنین logistic regression با L2 Regularization توسط شما باید انجام شود و مجاز به استفاده از کتابخانه هایی مانند sklearn نیستید. (بخش دوم قسمت ب، بخش سوم قسمت ب و بخش پنجم قسمت الف)

* بخش سوم قسمت ج، آموزش مدل با استفاده از گرادیان نزولی باید توسط شما پیاده سازی آن صورت گیرد و مجاز به استفاده از کتابخانه نیستید.

* در بقیه بخش های این سوال استفاده از کتابخانه دلخواه برای پیاده سازی آزاد است.

سوال ۷ (شبیه سازی، نمره ۱۵):

الف) بارگذاری و پیش‌پردازش: دیتاست Iris را بارگیری کرده و در یک dataframe ذخیره کنید. سپس ویژگی‌ها (X) و برچسب کلاس (y) را جدا کنید.

ب) مصورسازی: با استفاده از pairplot، رابطه بین ویژگی‌ها را با رنگ‌بندی بر اساس کلاس نمایش دهید. همچنین همبستگی بین ویژگی‌ها را با heatmap نمایش دهید و توضیح دهید کدام دو ویژگی بیشترین همبستگی را دارند.

ج) افزودن نویز: برای بررسی مقاومت مدل در شرایط غیرایده‌آل، با استفاده از توزیع نرمال (با میانگین=۰ و انحراف معیار=۱)، نویز به داده‌ها اضافه کنید. (مراحل بعدی را با داده‌های نویزدار انجام دهید.)

د) پیاده‌سازی SVM: برای هر یک از کرنل‌های زیر، مراحل ذکر شده را انجام دهید:

- Polynomial Kernel Function (degree = ۵)
- Gaussian RBF Kernel Function
- Sigmoid Kernel Function
- Linear Kernel Function

د۱. توضیح دهید که آن کرنل برای چه نوع داده‌ای مناسب‌تر است و بطور خلاصه معادله و پارامترهای آن را توضیح دهید.

د۲. با استفاده از هر دو استراتژی OVR و OVO، مدل را آموزش دهید. (این کار را حتما به کمک کلاس‌های OneVsRestClassifier و OneVsOneClassifier انجام دهید.)

۳.د Confusion Matrix و Classification Report را برای هر دو استراتژی نمایش دهید و نتایج OVR و OVO را از نظر دقت مقایسه کنید.

ه) نتیجه گیری: کدام کرنل بهترین و کدام کرنل بدترین عملکرد را داشت؟ پاسخ خود را با ذکر تحلیل مناسب بیان کنید.

* استفاده از کتابخانه های مختلف مانند sklearn و ... در این سوال مجاز است.

سوال ۸ (**شبیه سازی**، نمره ۱۰):

با توجه به مجموعه داده Q8-diabetes قرار داده شده پیاده سازی های زیر را انجام دهید.

الف) ابتدا داده های تست را ۲۰ درصد داده اصلی به صورت تصادفی و بدون جایگزینی انتخاب کنید و بقیه داده را به عنوان داده آموزش در نظر بگیرد.

ب) با استفاده از DecisionTreeClassifier و دادگان ضمیمه شده و با استفاده از معیار Gini درخت تصمیم را با مقدار پارامتر max_depth برابر با ۵ آموزش دهید.

ج) درخت آموزش داده شده را رسم کنید.

د) دقت این درخت آموزش دیده شده را روی داده تست را حساب کنید.

ه) دو ویژگی مهم که تاثیر زیادی در این دسته بندی داشتند را از روی این درخت تصمیم پیدا کنید.

و) صرفا دو ویژگی به دست امده در قسمت قبل را در نظر بگیرید و درخت تصمیم را مجدد با مقدار پارامتر max_depth برابر با ۵ آموزش دهید و مرز های تصمیم گیری را در این مدل را رسم کنید.

* در این سوال برای پیاده سازی ها از کتابخانه sklearn استفاده کنید و در همه بخش های کد مقدار random state را با عدد ۴۲ قرار دهید.