



دانشکده مهندسی برق و کامپیوتر

**درس**

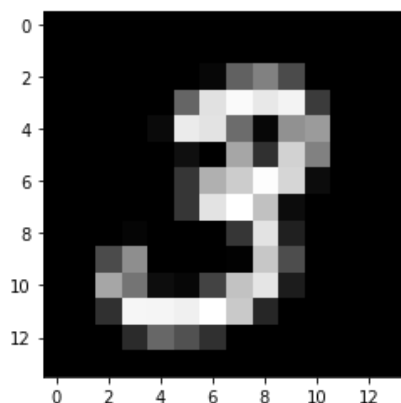
**یادگیری ماشین**

**تمرین کامپیوتری دوم**

نیمسال دوم سال تحصیلی ۱۴۰۰ - ۱۴۰۱

## سوالات

سوال ۱) در قسمت الف این سوال مجاز به استفاده از هیچ پکیج آماده ی یادگیری ماشین نیستید. هم چنین توصیه می شود برای آشنایی بیشتر با مجموعه داده ی Tiny MNIST ، بعضی نمونه های آن را رسم کنید و با ویژگی ها و ساختارهای آن ها آشنا شوید. به عنوان مثال من در شکل زیر اولین نمونه ی این مجموعه داده را رسم کرده ام. (۳۰ نمره)



(الف)

- ابتدا توضیح مختصری راجع به طبقه بندی های naïve bayes و optimal (non naïve) bayes داده و آن ها را با فرض گاوسی بودن داده های ورودی پیاده سازی کنید. دقت کنید که الگوریتم شما نباید هیچ پیش فرضی از داده های ورودی (اندازه ی آن ها و ...) داشته باشد و باید برای هر داده ای کار کند.
- الگوریتم هایی که پیاده سازی کرده اید را برای مجموعه داده ی Tiny MNIST (که نسخه ی فشرده ای از مجموعه داده ی MNIST است) تست کنید. نتیجه ی به دست آمده (شامل دقت طبقه بندی و f1 score برای هر کلاس) ذکر و راجع به تفاوت نتیجه ی دو روش بحث کنید.

(ب)

- از پکیج های آماده ی یادگیری ماشین استفاده کرده و مجموعه داده ی بالا را توسط الگوریتم naïve bayes در این پکیج طبقه بندی کرده و نتیجه را گزارش دهید.



سوال ۲) در این سوال بر روی مجموعه داده ی iris کار خواهید کرد و محدودیت نیز بر استفاده از هیچ پکیجی ندارید. مجموعه داده ی iris در پیوست قرار داده شده است. (۴۰ نمره)

با استفاده از الگوریتم SVM و روش های زیر، برای هر کدام، ناحیه ی کلاس های مختلف را رسم کنید. از ویژگی های Petal Length و Petal Width استفاده کنید.

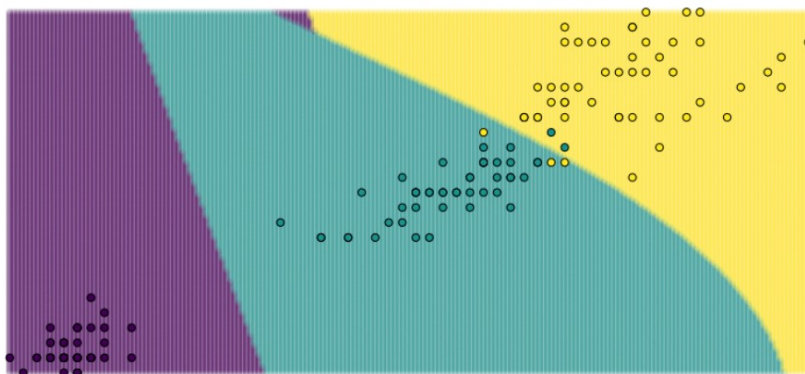
• SVM with linear kernel, one-vs-rest

• SVM with linear kernel, one-vs-one

• SVM with rbf kernel, one-vs-rest

• SVM with polynomial kernel ( $d=3$ ), one-vs-rest

علاوه بر رسم ناحیه های مختلف هر کلاس، برای هر کدام مقادیر دقت ترین و ماتریس آشفتگی را نیز گزارش کنید. شکل های ضمیمه شده ی شما، باید چیزی شبیه به شکل زیر باشد.





سوال ۳) در این سوال محدودیتی در استفاده از پکیج‌های یادگیری ماشین وجود ندارد. در این سوال از شما خواسته شده است تا اقدام به طراحی و اجرا و آزمون یک Decision Tree Classification کنید. بدین منظور دیتاست Social\_Network\_Ads.csv در اختیار شما قرار داده شده است که می‌توانید قسمتی از آن را در شکل زیر مشاهده نمایید: (۳۰ نمره)

Social\_Network\_Ads

Age	EstimatedSalary	Purchased
19	19000	0
35	20000	0
26	43000	0
27	57000	0
19	76000	0
27	58000	0
27	84000	0
32	150000	1
25	33000	0
35	65000	0
26	80000	0
26	52000	0
20	86000	0
32	18000	0

شکل ۱: قسمتی از دیتاست Social\_Network\_Ads.csv

در این دیتاست دو ویژگی Estimated Salary و Age به شما داده شده است و با استفاده از این دو باید پیش‌بینی کنید که با دیدن یک تبلیغ آیا شخص اقدام به خرید آن جنس می‌کند یا خیر.

الف. اقدام به لود کردن دیتاست خود و تقسیم کردن آن به دو مجموعه آموزشی و تست کنید. و در صورت نیاز بر روی دیتاست پیش پردازش نیز انجام دهید و اگر نیاز نیست علت آن را بیان کنید.

ب. یک درخت تصمیم ساخته و اقدام به آموزش آن با داده‌های آموزشی نمایید.

ج. دقت و ماتریس Confusion مربوط به این Classifier را به ازای داده‌های Train و Test، نمایش دهید.



د. فضای بدست آمده مجموعه‌های Train و Test مانند موردی که در شکل زیر نمایش داده شده است را نشان دهید.



شکل ۲: فضای نمونه داده‌های آموزشی دیتاست و دسته‌بندی آن‌ها با استفاده از Decision Tree

ه. درخت تصمیم حاصل را رسم نمایید.



## نکات پیاده سازی و تحویل.

- انجام این تمرین به صورت یک نفره می باشد.
- گزارش شما در فرآیند تصحیح از اهمیت ویژه‌ای برخوردار است. لطفاً تمامی نکات و فرضیهایی که برای پیاده‌سازی‌ها و محاسبات خود در نظر می‌گیرید را در گزارش ذکر کنید.
- از نوشتار محاوره‌ای و درج تصاویر دست نوشته در گزارش خودداری نمائید چون باعث کسر نمره خواهد شد.
- برای انجام تمرینها استفاده از زبان برنامه‌نویسی پایتون الزامی می‌باشد.
- در تمرین های برنامه نویسی حتما پیاده سازی خود را در محیط Jupyter Notebook و در یک فایل ipynb انجام دهید.
- نیازی به یک فایل پی دی اف جداگانه برای گزارش بخش پیاده‌سازی نیست. توضیحات خود را در همان فایل ipynb بنویسید. توضیحات به فارسی نوشته شوند در صورت تحویل فایل جداگانه‌ای برای گزارش نمره‌ی این بخش اعمال نخواهد شد.
- در فایل Jupyter Notebook هر سوال از تمرین به همراه پاسخ آن مشخص شده و خروجی‌های مورد نیاز نیز ذخیره شده باشد. همچنین هرگونه نتیجه و یا تحلیلی که در شرح سوال از شما خواسته شده است را به طور واضح و کامل در گزارش بیاورید. در صورت عدم رعایت این مورد، بدیهی است که از نمره تمرین کسر می‌شود.
- بعد از تکمیل پاسخ‌ها در فایل Jupyter Notebook مجدداً kernel را راه‌اندازی کرده و فایل را اجرا نمایید به صورتیکه شماره‌ی هر سلول در فایل دقیقاً مطابق با ترتیب سلول‌ها باشد.
- تکالیف کامپیوتری تا یک هفته بعد از موعد مقرر قابل تحویل می‌باشند و به ازاء هر روز تأخیر ۷٪ از نمره کل کسر می‌گردد.
- در صورت مشاهده تقلب امتیاز تمامی افراد شرکت‌کننده در آن، • لحاظ می‌شود.
- در صورت وجود سوال و یا ابهام میتوانید از طریق آیدی تلگرام زیر با دستیار آموزشی در ارتباط باشید:

@givkashi

@basir\_ebr

@hamidravace