



پردیس دانشکده های فنی

به نام خدا
دانشکده‌ی مهندسی برق و کامپیوتر
تمرین سری سوم یادگیری ماشین



دانشگاه تهران

سلام بر دانشجویان عزیز، چند نکته مهم:

1. حجم گزارش به هیچ عنوان معیار نمره دهی نیست، در حد نیاز توضیح دهید.
2. نکته‌ی مهم در گزارش نویسی روشن بودن پاسخ‌ها می‌باشد، اگر فرضی برای حل سوال استفاده می‌کنید حتماً آن را ذکر کنید، اگر جواب نهایی عددی است به صورت واضح آن را بیان کنید.
3. کدهای ارسال شده بدون گزارش فاقد نمره می‌باشند.
4. برای سوالات شبیه سازی، فقط از دیتاست داده شده استفاده کنید.
5. فایل نهایی خود را در یک فایل زیپ شامل، pdf گزارش و فایل کدها آپلود کنید. نام فایل زیپ ارسالی الگوی ML_HW3_StudentNumber داشته باشد.
6. نمره تمرین ۱۰۰ نمره می‌باشد.
7. هرگونه شباهت در گزارش و کد مربوط به شبیه سازی، به منزله تقلب می‌باشد و کل تمرین برای طرفین **صفر** خواهد شد.
8. در صورت داشتن سوال، از طریق ایمیل ershad.hasanpour@ut.ac.ir سوال خود را مطرح کنید.

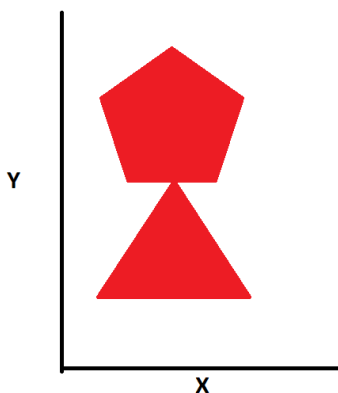
سوال ۱: (۱۲ نمره)

- در این سوال با مفاهیم اساسی شبکه های اساسی سر و کار دارید و باید به پرسش های زیر پاسخ دهید.
- الف) نقش توابع هزینه و همچنین توابع فعالساز را در شبکه های عصبی توضیح دهید.
- ب) آیا مقدار دهی اولیه وزن ها و بایاس ها در یادگیری شبکه های عصبی تاثیر دارد؟ توضیح دهید.
- ج) تفاوت بین **Forward Propagation** و **Backward Propagation** را توضیح دهید.
- د) مقدار نرخ یادگیری (بزرگ بودن یا کوچک بودن آن) ، چگونه می تواند آموزش یک شبکه عصبی را تحت تاثیر قرار دهد.
- ه) توضیح دهید چرا شبکه های عصبی با لایه های بیشتر بهتر از شبکه های با لایه های کمتر می باشند؟
- و) بزرگترین مشکل شبکه های عصبی در مرحله آموزش چیست؟ این مشکل را توضیح دهید و سپس یک راه حل برای آن پیشنهاد دهید.

سوال ۲: (۹ نمره)

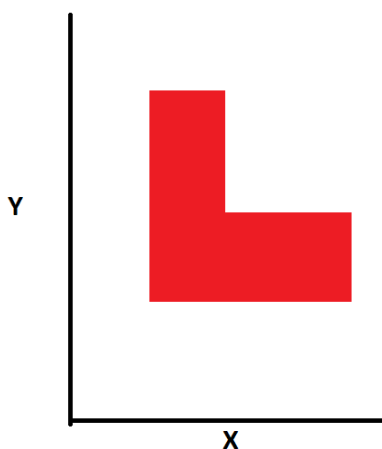
در این سوال با معماری شبکه های عصبی سروکار دارید. در واقع باید برای طبقه بندی هر یک از موارد زیر، یک شبکه عصبی MLP پیشنهاد دهید. لازم به ذکر است که صرفاً تعداد لایه ها و تعداد نورون های هر لایه را در معماری هایتان لحاظ کنید و نیازی به یادگیری وزنها ندارید.

الف) معماری پیشنهادی تان باید بتواند ناحیه قرمز رنگ موجود در شکل (۱) را از سایر فضای دوبعدی جداسازی کند.



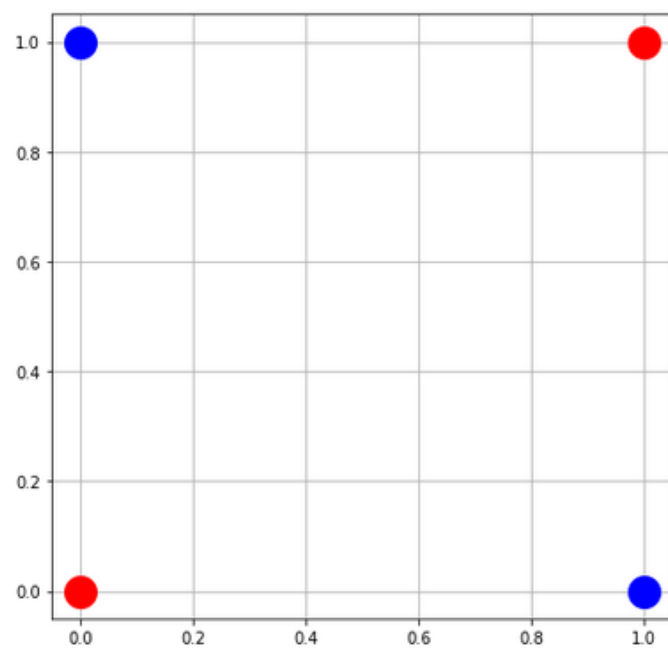
شکل (۱)

ب) معماری پیشنهادی تان باید بتواند ناحیه قرمز رنگ موجود در شکل (۲) را از سایر فضای دوبعدی جداسازی کند.



شکل (۲)

ج) معماری پیشنهادی تان باید بتواند داده های آبی شکل (۳) را از داده های قرمز جدا کند.

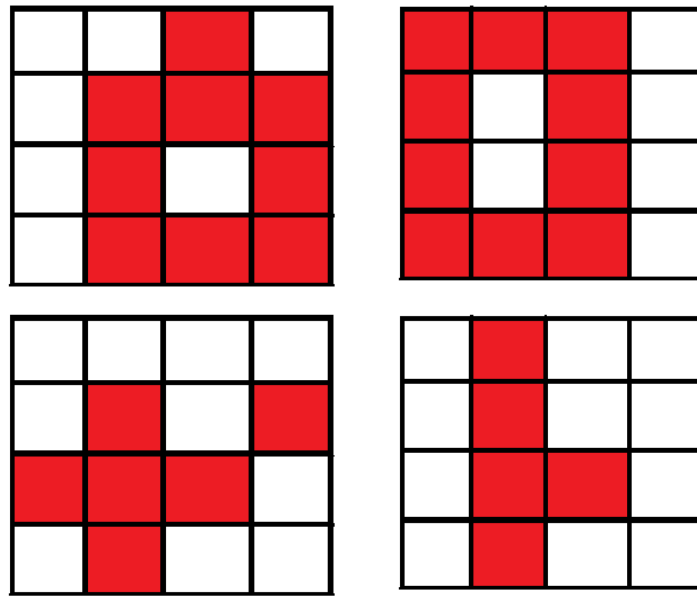


شكل (3)

سوال ۳: (۹ نمره)

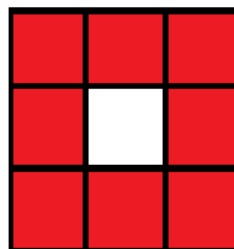
به سوالات زیر پاسخ دهید.

الف) چهار تصویر 4×4 زیر را در نظر داشته باشید. تصاویر ردیف اول مربوط به کلاس 1 و تصاویر ردیف دوم مربوط به کلاس دوم می باشند. پیکسل های قرمز رنگ عدد 1 و پیکسل های سفید رنگ عدد صفر را نشان می دهند. مشخصه بارزی که تصاویر کلاس اول نسبت به کلاس دوم دارند چیست؟



شکل (4)

ب) فیلتر زیر را در نظر بگیرید. عملیات کانولوشن این فیلتر روی هر یک از این تصاویر را انجام دهید. فرض کنید عدد بایاس این فیلتر برابر 2- بوده و همچنین فرض کنید که از تابع فعالساز RELU استفاده شده است. پس از اعمال کانولوشن روی هر تصویر، یک فیچر 2×2 از آن تصویر بدست می آید. حال این فیچر بدست آمده را نیز از یک maxpooling به اندازه 2×2 عبور دهید تا در خروجی هر تصویر به یک عدد واحد برسید. به عبارتی با این عملیات هر عدد خروجی معرف یک تصویر خواهد بود.



شکل (۵)

ج) با توجه به خروجی های بدست آمده در بخش قبلی، توضیح دهید چگونه می توان با استفاده از این خروجی ها، تصاویر مربوط به دو کلاس را از هم جدا کرد.

سوال ۴: (۸ نمره)

به سوالات زیر که در مورد توابع فعالساز می باشند، پاسخ دهید.

الف) یکی از مزیت های اصلی تابع فعالساز $relu$ نسبت به توابع فعالساز $sigmoid$ و $tanh$ ، این می باشد که بار محاسباتی کمتری دارد. با توجه به ماهیت این تابع نشان دهید چرا تابع فعالساز $relu$ نسبت به دو تابع دیگر بار محاسباتی کمتری دارد.

ب) یکی دیگر از مزیت های تابع فعالساز $relu$ این می باشد که برخلاف دو تابع فعالساز $sigmoid$ و $tanh$ ، دچار مشکل $vanishing gradient$ نمی شود. این مشکل را توضیح داده و استدلال کنید چرا $sigmoid$ و $tanh$ دچار این مشکل می شوند ولی $relu$ دچار این مشکل نمی شود.

ج) در یک مساله طبقه بندی در لایه آخر از تابع فعالساز $relu$ استفاده خواهید کرد یا $sigmoid$ ؟ چرا؟

د) فرق بین تابع فعالساز $leaky relu$ با $relu$ در چیست؟ در چه مواردی بهتر است که به جای $relu$ از $leaky relu$ استفاده شود؟

سوال ۵: (۱۲ نمره)

در این سوال می خواهیم استخدام تعدادی از داوطلبان برای کار در یک شرکت هوش مصنوعی را بررسی کنیم. برای این کار می خواهیم از یک درخت تصمیم استفاده کنیم. فرض کنید داده های زیر را در اختیار داریم.

جدول 1: درخت تصمیم

استخدام	تحصیلات دانشگاهی	جنسیت	سابقه کاری برحسب سال
رد	کارشناسی	مرد	1
قبول	کارشناسی	زن	2
رد	دکتری	مرد	0
قبول	کارشناسی	مرد	2
رد	کارشناسی ارشد	مرد	1
رد	دکتری	زن	1
قبول	کارشناسی ارشد	زن	0
قبول	کارشناسی	زن	1
رد	کارشناسی ارشد	مرد	0
رد	کارشناسی ارشد	مرد	2

الف) آنتروپی استخدام را حساب کنید.

ب) میزان **information gain** را برای ویژگی "تحصیلات دانشگاهی" و همچنین برای ویژگی "سابقه کاری" بدست آورید.

ج) اگر یک درخت تصمیم با ارتفاع 1 با استفاده از ویژگی "سابقه کاری" بسازیم، مقادیر **recall** ، **accuracy** ، **precision** و **specificity** را بدست آورید.

د) قسمت ج را برای وقتی که درخت تصمیم مان با ارتفاع 1 و با استفاده از ویژگی "جنسیت" ساخته شود، تکرار کنید.

سوال ۶: (شبیه‌سازی، ۱۵ نمره)

در این سوال قصد داریم به طبقه بندی سیگنال های ECG بپردازیم. در ابتدا کار استخراج ویژگی روی سیگنال های ECG انجام شده است و برای هر سیگنال 169 تا ویژگی مختلف بدست آمده است. بنابراین دیتاستی که با آن کار خواهید کرد، سیگنال های خام ECG نیستند بلکه فیچر هایی می باشند که از این سیگنال های خام استخراج شده اند. این دیتاست شامل داده های ECG برای چهار کلاس مختلف می باشد. کلاس اول مربوط به داده های افراد نرمال (با علامت N) ، کلاس دوم مربوط داده های فیبریلاسیون دهلیزی (با علامت A) ، کلاس سوم مربوط به داده های سایر بیماری ها (با علامت O) و کلاس چهارم مربوط به داده هایی می باشند که در حین ثبت، نویز نامشخصی به آنها اضافه شده است (با علامت ~).

الف) تعداد داده های مربوط به هریک از این چهار کلاس را با استفاده از کتابخانه Pandas بدست آورید.

ب) توزیع داده ها را بر اساس کلاسه‌شان به صورت هیستوگرام نشان دهید. با توجه به هیستوگرام مربوطه توضیح دهید که این توزیع نامتوازن چه مشکلی را می تواند در مساله طبقه بندی به همراه داشته باشد.

ج) یک شبکه MLP با 4 لایه و همچنین تعداد نورون های مناسب طراحی کنید. این شبکه را روی دیتاست مربوطه پیاده سازی کنید. دیتاست مربوطه را با نسبت 0.9 به 0.1 به داده های ترین و تست تقسیم کنید. مقادیر precision ، recall ، f1-score را برای هر کلاس، مقدار دقت کل و همچنین ماتریس کانفیوژن را بدست آورده و گزارش کنید. این شبکه داده های مربوط به کدام کلاس را بهتر از سایر کلاسه‌ها طبقه بندی می کند و چرا؟

د) حال دیتاست مربوطه را با استفاده از روش استاندارد^۱، نرمالیزه کنید. تمامی خواسته های بخش ج را برای دیتاست نرمالیزه شده تکرار کنید.

ه) نتایج بدست آمده در بخش ج و د را با هم مقایسه کرده و تحلیل کنید که چرا نتایج مربوط به بخش د بهتر از نتایج بخش ج می باشند.

و) در دیتاست داده شده داده های مربوط به کلاس چهارم (کلاسی که کمترین تعداد داده را دارد) را حذف کنید و داده های مربوط به کلاس دوم و سوم را نیز با هم ترکیب کنید و یک دیتاست دو کلاسه normal و abnormal تهیه کنید. حال بخش د را برای این دیتاست دو کلاسه تکرار کنید.

¹ StandardScaler

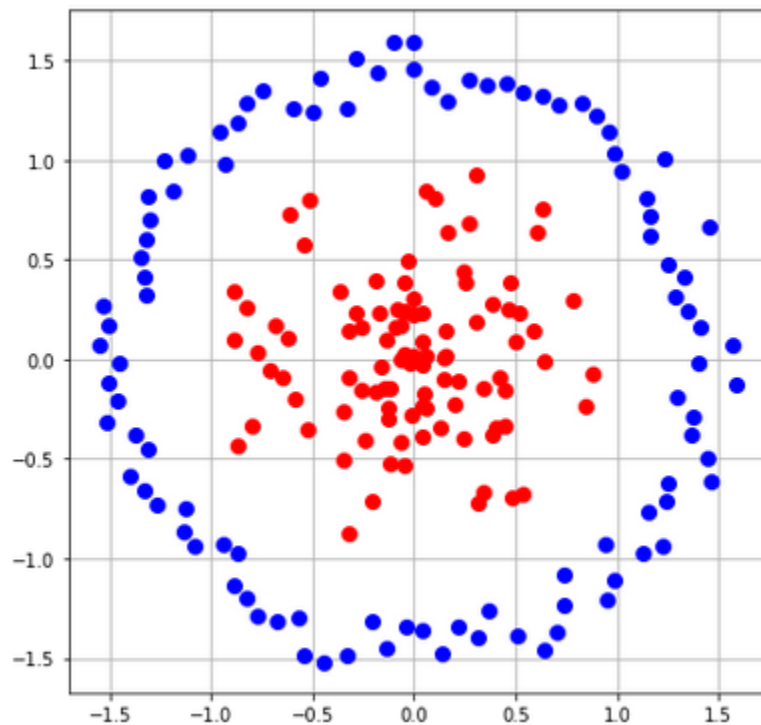
سوال ۷: (شبیه‌سازی، ۱۵ نمره)

الف) یک دیتاست دو کلاسه و دو بعدی به صورت زیر تولید کنید.

کلاس اول: شامل 100 نقطه درون دایره ای به مرکز $(0,0)$ و محدود به شعاع های 0 و 1

کلاس دوم: شامل 100 نقطه درون دایره ای به مرکز $(0,0)$ و محدود به شعاع های 1.3 و 1.6

توزیع داده های تان باید چیزی مشابه با شکل 1 باشد.



شکل 1: توزیع داده های دو کلاس

ب) با استفاده از یک شبکه MADALINE ، با 3 نورون در لایه آخر ، داده های قرمز رنگ را از داده های آبی جداسازی کنید. لازم به ذکر است که در این سوال شما نمی توانید از کتابخانه های آماده استفاده کنید. خطوط جداساز بدست آمده را روی نمودار رسم کنید و با توجه به خطوط بدست آمده دقت طبقه بندی را بدست آورده و گزارش کنید.

ج) قسمت ب را با یک شبکه MADALINE با 6 نورون در لایه آخر تکرار کنید.

د) نتایج قسمت ب و ج را با هم مقایسه کرده و تحلیل کنید.

سوال ۸: (شبیه‌سازی، ۱۰ نمره)

در این سوال به پیاده سازی یک شبکه کانولوشنی برای طبقه بندی تصاویر cifar10، خواهید پرداخت. این دیتاست 10 کلاسه بوده و در مجموع شامل 60000 تصویر می باشد.

الف) با استفاده از قطعه کد زیر این دیتاست را دانلود کرده و 5 تصویر را به صورت رندوم نمایش دهید.

```
from keras.datasets import cifar10
```

ب) پیش پردازش های لازم را روی این داده ها انجام داده و آنها را به سه گروه تست و ترین و ولیدیشن تقسیم کنید. (با نسبت 40000/10000/10000)

ج) یک شبکه کانولوشنی عمیق با معماری زیر را طراحی کنید تا این کلاس ها را از هم تفکیک کند. این معماری را با solver های SGD ، Adam و rmsprop آموزش دهید و برای هر کدام از آنها نمودار دقت و خطا را برای داده های ترین و ولیدیشن نمایش دهید.

جدول 2: معماری مربوط به شبکه CNN

Layers	Layers Parameter	Activation Function
Conv2D	32 , size(3,3)	Relu
Conv2D	32 , size(3,3)	Relu
Conv2D	32 , size(3,3)	Relu
Maxpooling2D	Size(2,2)	-
Droupout	0.25	-
Conv2D	64 , size(3,3)	Relu
Conv2D	64 , size(3,3)	Relu
Conv2D	64 , size(3,3)	Relu
Maxpooling2D	Size(2,2)	-
Droupout	0.25	-
Dense	512	Relu
Droupout	0.5	-
Dense	10	Softmax

د) برای هر یک از این شبکه ها، مقادیر دقت، precision ، recall و f1-score را برای داده های تست گزارش کنید.

سوال ۹: (شبیه‌سازی، ۱۰ نمره)

- در این سوال با دیتاست Diabetes.csv برای طبقه بندی افراد دیابتی و غیر دیابتی، کار خواهید کرد.
- الف) در ابتدا داده ها را با نسبت 7 به 3 به داده های ترین و تست تقسیم کرده و یک درخت تصمیم را روی آنها اعمال کنید. دقت را برای داده های تست و ترین گزارش کرده و با هم مقایسه کنید.
- ب) با استفاده از کتابخانه های آماده، درخت تصمیمی که ایجاد کرده اید را نمایش دهید.
- ج) همان طور که می دانید یکی از روش های بهتر کردن نتایج مربوط به درخت تصمیم، pre-pruning می باشد. این روش را مختصرا توضیح دهید.
- د) حال درختی را طراحی کنید که حداکثر عمق آن برابر 2 باشد. مقادیر دقت را برای داده های تست و ترین گزارش کرده و با نتایج بدست آمده در قسمت الف مقایسه کنید.
- ه) درخت مربوط به بخش د را نمایش داده و آن را از لحاظ تفسیرپذیری با نتایج بخش ب مقایسه کنید.