

سلام بر دانشجویان عزیز، چند نکته مهم:

۱. حجم گزارش به هیچ عنوان معیار نمره دهی نیست، در حد نیاز توضیح دهید.
۲. نکته ی مهم در گزارش نویسی روشن بودن پاسخ ها می باشد، اگر فرضی برای حل سوال استفاده می کنید حتما آن را ذکر کنید، اگر جواب نهایی عددی است به صورت واضح آن را بیان کنید.
۳. کدهای ارسال شده بدون گزارش فاقد نمره می باشند.
۴. 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.
۵. نمره تمرین ۱۰۰ نمره می باشد و حداکثر تا نمره ۱۱۰ (۱۰ نمره امتیازی) می توانید کسب کنید.
۶. هرگونه شباهت در گزارش و کد مربوط به شبیه سازی، به منزله تقلب می باشد و کل تمرین برای طرفین صفر خواهد شد.
۷. در صورت داشتن سوال، از طریق ایمیل های زیر سوال خود را مطرح کنید.

سوالات 1 و 2 و 3: smousavichashmi@ut.ac.ir

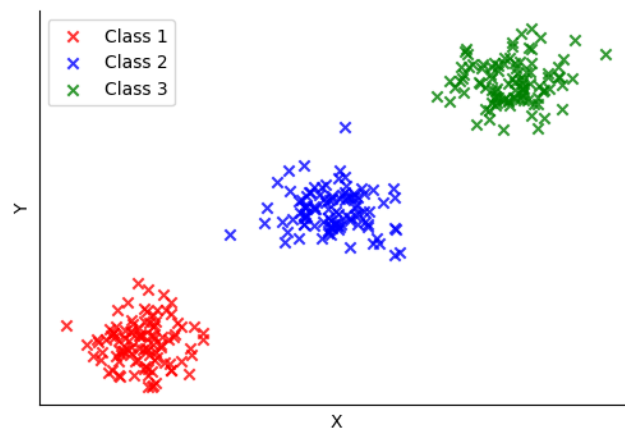
سوالات 4, 5 و 6: navidrazzaghi1381@gmail.com

سوال ۱: (۱۵ نمره)

الف) هدف توابع فعال ساز در شبکه های MLP چیست و اگر در شبکه MLP از توابع فعال ساز استفاده نکنیم چه مشکلی ایجاد می شود؟

ب) مشکل رایج تابع فعال سازی Sigmoid چیست و چه تابع فعال سازی و به چه صورت می تواند آن مشکل را برطرف کند.

ج) برای دسته بندی داده های کلاس های زیر یک شبکه MLP به گونه ای طراحی کنید که دارای کمترین تعداد لایه و تعداد نرون در هر لایه باشد.



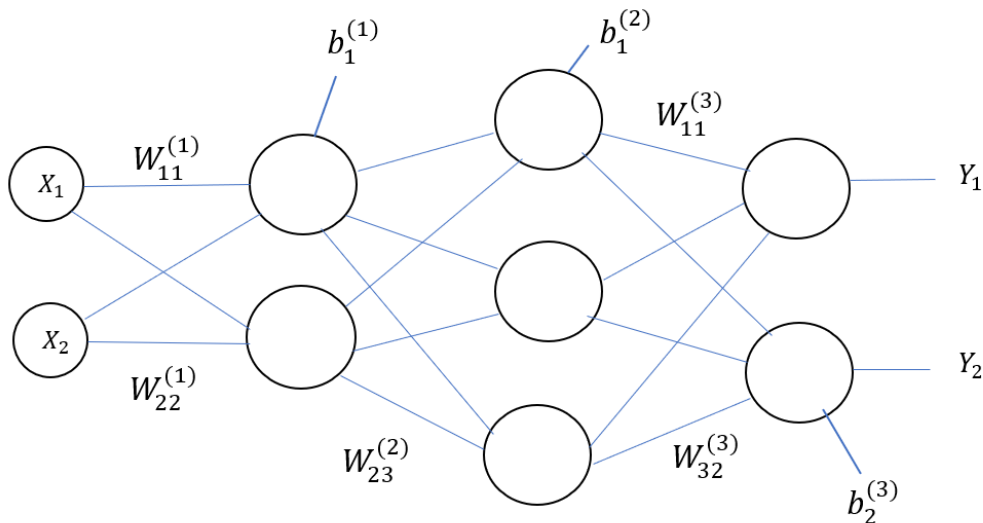
د) یک شبکه MLP را با استفاده از تابع فعال سازی ReLU به نحوی طراحی کنید که معادل تابع زیر باشد.

$$T(x) = \begin{cases} 3x & \text{if } 0 \leq x < \frac{1}{3} \\ \frac{3}{2}(1-x) & \text{if } \frac{1}{3} \leq x \leq 1 \end{cases} \quad x \in \mathbb{R}$$

ه) با ذکر مثال توضیح دهید چرا یک مدل MLP می تواند یک Universal approximator برای هر تابعی باشد.

سوال ۲: (۱۵ نمره)

الف) شبکه MLP زیر را در نظر بگیرید که برای مسئله دسته بندی دو کلاسه است. در این شبکه تابع فعال سازی لایه آخر را Softmax و توابع فعال سازی دیگر لایه ها Sigmoid در نظر بگیرید.



با توجه به الگوریتم backpropagation گرادیان وزن های پارامتر های $W_{11}^{(1)}$ و $b_1^{(1)}$ که در شکل مشخص شده است) نسبت خطای cross entropy بر اساس تک داده ورودی با ذکر راه حل به صورت پارامتری بدست آورید. (نحوه نام گذاری پارامتر ها بر اساس نام گذاری های داده شده در شکل قابل استنباط است)

ب) در مورد الگوریتم های بهینه سازی مبتنی بر گرادیان زیر تحقیق کنید و نحوه کارکرد آن ها را توضیح دهید.

۱- Stochastic gradient descent

۲- Newton-Raphson method

سوال ۳: (شبیه سازی، ۲۰ نمره)

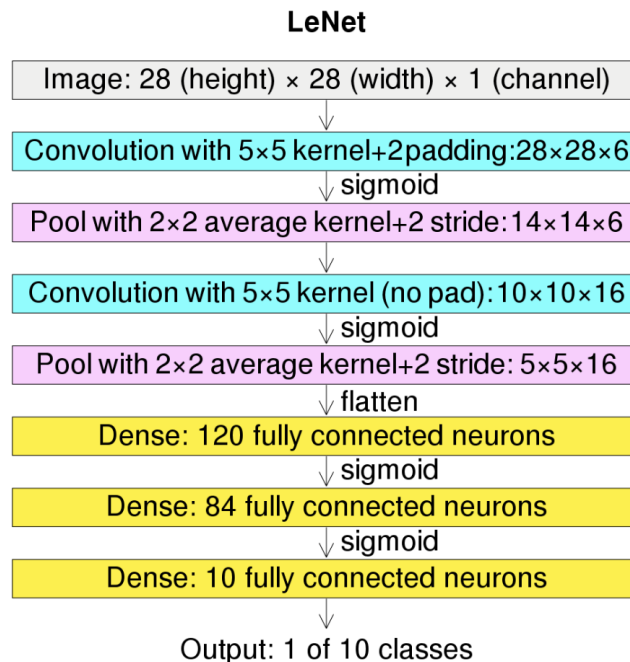
الف) یکی از ویژگی های شبکه های عصبی کانولوشنی، Transitional invariance بودن این شبکه ها است. ویژگی Transitional invariance بودن شبکه های عصبی کانولوشنی به چه معنا است؟

ب) به طور دقیق توضیح دهید چه اجزایی از شبکه عصبی کانولوشنی باعث ایجاد ویژگی Transitional invariance می شود.

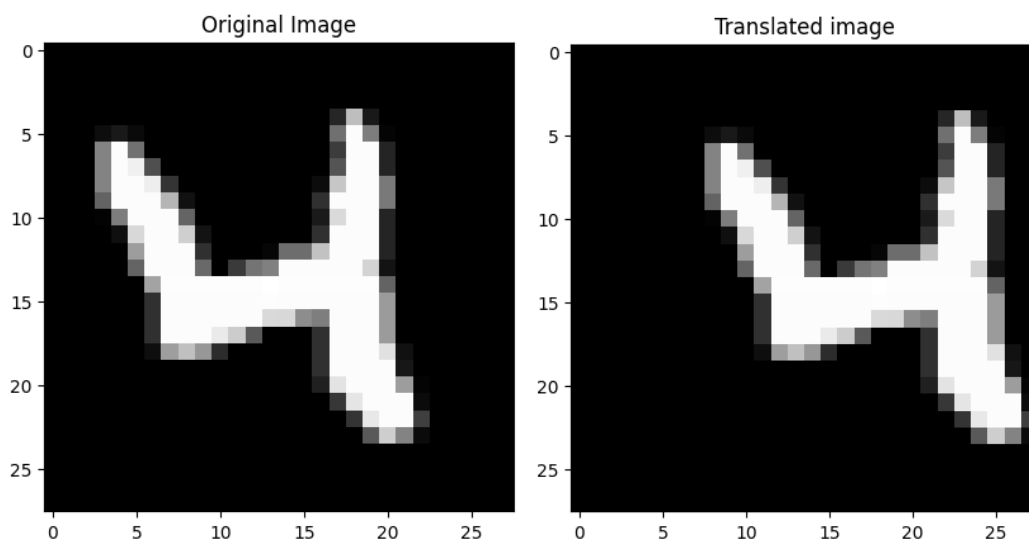
ج) در ادامه می خواهیم به بررسی ویژگی Transitional invariance در شبکه های CNN بپردازیم و آن را با شبکه های MLP مقایسه کنیم.

1- در ابتدا یک شبکه MLP با معماری دلخواه طراحی کنید و روی مجموعه آموزش مجموعه داده MNIST آموزش دهید. توجه کنید که شبکه طراحی شده شما حداقل به دقت 90 درصد روی داده تست مجموعه داده MNIST برسد.

2- سپس شبکه LeNet را که معماری آن در شکل زیر نشان داده شده است را پیاده سازی کنید و روی مجموعه داده MNIST آموزش دهید. انتظار می رود که به دقت حداقل بالا 95 درصد روی داده تست مجموعه داده MNIST برسید.

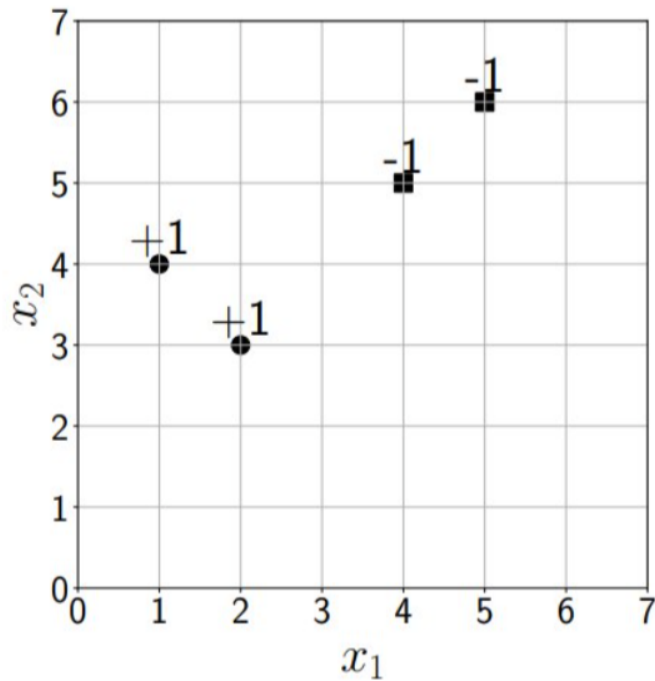


3- سپس یک تصویر از مجموعه داده MNIST در نظر بگیرید و عکس را چند پیکسل شیفت دهید با این هدف که عدد عکس درون آن جا به جا شود. توجه کنید این جا به جایی نباید به صورتی باشد که عدد از کادر درون تصویر خارج شود. سپس این عکس را به شبکه های MLP و CNN آموزش داده شده بدهید و کلاس های پیش بینی شده برای آن را مقایسه کنید. در صورتی که به نتیجه مورد نظر نرسیدید عکس با کلاس های متعدد و شدت اعمال Translation از نظر دوری و نزدیکی امتحان کنید (در پیاده سازی های این تمرین می توانید از کتابخانه های آماده استفاده کنید) .



سوال ۴: (شبیه سازی، ۲۰ نمره)

الف) پایگاه داده زیر را در نظر بگیرید. فرض کنید یک ماشین بردار پشتیبان (Hard SVM) بر روی این پایگاه داده آموزش داده ایم. مقدار بایاس b و بردار وزن $[w_1, w_2]^T$ را بدست آورید. صحت محاسبات خود را با استفاده از توابع آماده مربوط به SVM در پایتون بررسی کنید.



ب) هر کدام از پایگاه داده های زیر شامل نقاطی است که متعلق به کلاس $1+$ یا $1-$ است. برای هر کدام از پایگاه داده ها، یک نگاشتی پیدا کنید که نقاط در فضای تبدیل یافته، به صورت خطی جدایی پذیر باشند.

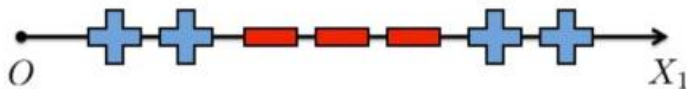


Figure 1 پایگاه داده اول

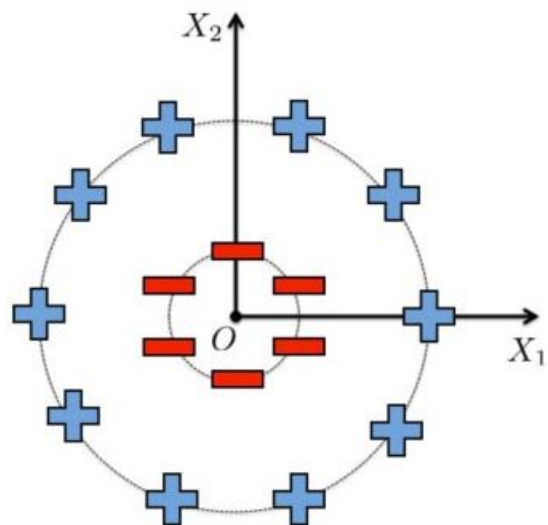


Figure 2 پایگاه داده دوم

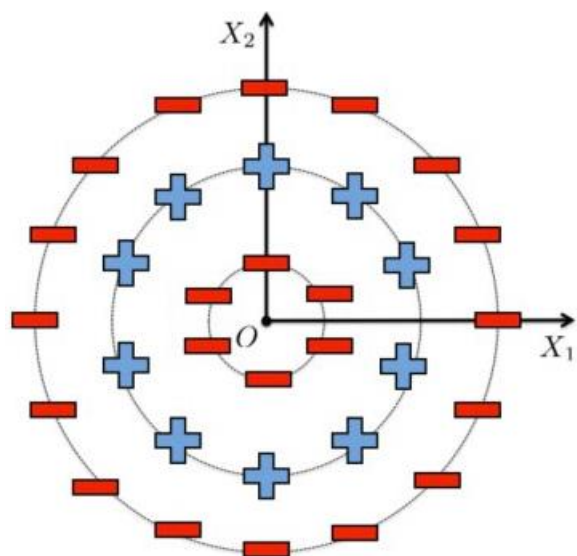


Figure 3 پایگاه داده سوم

سوال ۵: (۲۰ نمره)

مفهوم کلی کرنل و دلایل استفاده از روش‌های مبتنی بر کرنل را بیان کنید و برای یک کرنل معتبر، اثبات کنید که داریم

$$K(x, y)^2 \leq K(x, x)K(y, y)$$

هم‌چنین فرض کنید مجموعه داده‌ی $S = \{x_q\}_{q=1}^Q$ داده شده است. تحت تبدیل ϕ این نقاط از فضای d_1 بعدی به فضای d_2 بعدی انتقال پیدا کرده‌اند. اگر میانگین نقاط در این فضای جدید و کرنل مربوط با تبدیل ϕ را به ترتیب با μ_ϕ و K_ϕ نشان دهیم، اثبات کنید داریم

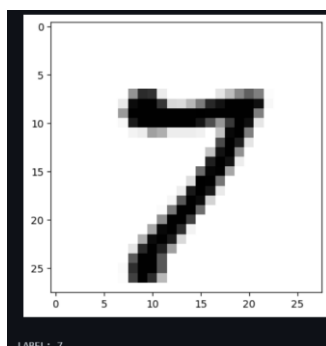
$$\|\mu_\phi\| = \frac{1}{Q} \sqrt{\sum_{m=1}^Q \sum_{n=1}^Q K_\phi(x_m, x_n)}$$

سوال ۶: (شبیه سازی، ۲۰ نمره)

در این سوال می خواهیم استفاده از ماشین های بردار پشتیبان (SVM) را برای طبقه بندی تصاویر بررسی کنیم. ما می خواهیم از مجموعه داده معروف MNIST استفاده کنیم، که مجموعه داده ای از ارقام دست نویس است. ما داده ها را از mldata.org دریافت می کنیم، که یک repository عمومی برای داده های یادگیری ماشین است.

مجموعه داده شامل 70000 تصویر از ارقام دست نویس است (یعنی 0، 1، ... 9). هر تصویر 28 پیکسل در 28 پیکسل است و می توانیم آن را بردار $28 \times 28 = 784$ عدد در نظر بگیریم. هر عدد یک عدد صحیح بین 0 تا 255 است. برای هر تصویر برچسب مربوطه را داریم (یعنی 0، 1، ... 9).
مراحل زیر را انجام دهید:

- پایگاه داده را به داده های آموزش و تست تقسیم بندی کنید . مطمئن باشید که از هر کدام از لیبل های موجود، حداقل 10 تا نمونه در مجموعه آموزشی موجود باشد
- تابعی تعریف کنید که یک تصویر از پایگاه داده را مانند مثال زیر چاپ کند :



- مدل SVM روی داده های آموزش، ترین کنید. سه کرنل linear, polynomial degree 2 و RBF را در مدل SVM لحاظ کنید. با استفاده از grid search بهترین پارامتر هایی که برای هر کرنل بدست می آید را گزارش کرده و برای هر کدام score را بیان کنید.

- با توجه به قسمت قبل، بهترین مدل SVM یعنی آن مدلی که بیشتری score را داشته است را گزارش کنید (به همراه ذکر پارامتر های آن مدل)
- بهترین مدل SVM را روی داده های آموزش ، ترین کرده و خطای آموزش را گزارش کنید. همچنین با پیش بینی لیبل داده های تست، خطای تست را گزارش کنید.
- برای مقایسه، از لاجیستیک رگرسیون استفاده کرده و خطای تست لاجیستیک رگرسیون و بهترین مدل SVM را مقایسه کنید.
- کدی بنویسید که رقمی را پیدا و نمایش دهد که توسط لاجیستیک رگرسیون به اشتباه و توسط بهترین مدل SVM درست طبقه بندی می شود.