



- هر گروه فقط باید یک فایل نوت بوک ارسال کند که در آن مشخصات اعضای گروه در ابتدای نوت بوک قرار داشته باشد.
- جهت اضافه شدن به گروه تلگرامی این پروژه و داشتن ارتباط مستقیم با دستیاران از لینک زیر استفاده کنید:
- **ML-FinalProject**
- در صورت هم فکری و یا استفاده از هر منابع خارج درسی، نام هم فکران و آدرس منابع مورد استفاده برای حل سوال مورد نظر را ذکر کنید.
- لطفا کدهای خود را به همراه توضیحات کاملا کاملت گذاری کنید و در نظر داشته باشید مرتب بودن کدها و خوانایی آن‌ها در ارزشیابی نمره هر سوال تاثیر دارد
- این پروژه تحویل و ارائه مجازی دارد که ده درصد از نمره پروژه به آن تعلق دارد و در ادامه زمان آن در تقویم درس مشخص می شود و تمام اعضای گروه باید به کدها و مطالب آن تسلط کامل داشته باشند.

## (سوالات)

### مسئله ۱. (۳۰ نمره)

در این سوال شما باید به حل مسئله تشخیص بیماری قلبی با استفاده از ماشین های بردار پشتیبان بپردازید. توضیحات سوال در فایل نوت بوک قرار داد و میتوانید به آن مراجعه کنید.

### مسئله ۲. (۳۰ نمره)

در این سوال شما الگوریتم Adaboost را با دسته بندی های پایه decision stump پیاده سازی می کنید تا عمل دسته بندی روی مجموعه داده ساختگی adaboost-synndata انجام دهید. یک decision stump یک جدا کننده خطی است که موازی یکی از محورهاست و نمونه های یک سویش را مثبت و نمونه های سوی دیگر را منفی دسته بندی می کند. مثلا فرض کنید ورودی ها دوی بعدی هستند. در این صورت از decision stump های زیر استفاده می کنیم:

$$h(x) = \begin{cases} y & x \geq k \\ -y & o.w \end{cases}$$

یا

$$h(x) = \begin{cases} y & x_1 \leq k \\ -y & o.w \end{cases}$$

که  $y \in \{1, -1\}$  و  $k$  یک عدد حقیقی دلخواه است. طبق فایل نوت بوک، ابتدا الگوریتم Adaboost را پیاده سازی کنید و و آموزش دهید. سپس به سوالات زیر پاسخ دهید.

الف) در هر دور اجرای الگوریتم، خطای وزن دار  $\epsilon_t$  دسته بند ضعیف خود  $h_t$  را روی داده های آموزش محاسبه کنید. برای  $T = 20$  نمودار  $\epsilon_t$  را بر حسب  $t$  رسم کنید. بیشترین خطای یک دسته بند چقدر است؟ کجا رخ می دهد؟ نمودار را تفسیر کنید.

ب) در هر دور نرخ خطای دسته بند نهایی روی داده های آموزش و تست  $H_t$  را محاسبه کنید. در  $T = 20$  نمودار این دو خطا را بر حسب  $t$ ، روی یکدیگر رسم کنید. نمودارها را تفسیر کنید.

### مسئله ۳. (۳۰ نمره)

در این مسئله قصد داریم تصاویر مجموعه دادگان Fashion-MNIST را دسته بندی کنیم. برای دریافت مجموعه دادگان میتوانید از یکی از کتابخانه های موجود استفاده کنید.

الف) ابتدا تابعی بنویسید که با هر بار اجرا ۱۶ تصویر از این مجموعه دادگان را بصورت تصادفی در یک گرید ۴ در ۴ نمایش می دهد و آن را یک بار اجرا کنید.

ب) یک شبکه MLP با سه لایه نهان با اندازه های ۱۲۸ و ۶۴ و ۶۴ پیاده سازی کنید. برای اینکار در ابتدا تنها مجاز هستید از کتابخانه Pytorch استفاده کنید. در اینجا نیاز دارید تا تمامی مراحل Backpropagation را پیاده سازی کنید و وزن ها را بر اساس گرادیان های محاسبه شده بروزرسانی کنید، انتخاب الگوریتم بهینه سازی مناسب و تابع هزینه مناسب بسته به نظر شما دارد که در زمان تحویل پروژه باید آن را توضیح دهید.

ج) نمودار تابع هزینه در هنگام آموزش را رسم کنید

د) دقت دسته بندی بر روی داده های آزمون را گزارش کنید

موفق باشید (:)