

تمرین عملی سری دوم درس مقدمه‌ای بر یادگیری ماشین: شبکه عصبی

سوال اول: هدف این تمرین پیش بینی هزینه اقامت در هتل به ازای ویژگی های (ستون های) مختلفی میباشد که مجموعه دادگان آن در فایل `Hotel.zip` آپلود شده است. از داده های فایل `H1.csv` به عنوان داده های `train` و `validation` و از داده های فایل `H2.csv` به عنوان داده های `test` استفاده نمایید.

الف) در این قسمت می بایست از تمامی ویژگی ها (`categorical` , `numerical`) به عنوان `feature` استفاده نمایید و شبکه عصبی چند لایه ای را طراحی کنید که هزینه اقامت در هتل را به ازای تمامی این `feature` ها پیش بینی نماید. نسبت داده های `train` به `validation` می بایست ۸۰ به ۲۰ باشد. نمودار `loss` را به ازای ۳۰، ۵۰ و ۱۰۰ اپیاک رسم نمایید.

نکته: می بایست در ابتدا ویژگی های `numerical` و `categorical` را از هم تشخیص داده و برای هر کدام نرمال سازی های لازم را انجام دهید تا بتوانید از این داده ها در شبکه بهره ببرید.

نکته : میتوانید از کتابخانه `pandas` به منظور نرمال سازی ویژگی های `categorical` بهره ببرید.

نکته: می بایست ویژگی `Average Daily Rates` را پیش بینی نمایید. (هزینه اقامت هتل) و به غیر از آن ۳۴ ویژگی دیگر وجود دارد که بر اساس آن ها `ADR` پیش بینی میشود.

ب) ارزیابی را برحسب معیار های `MSE` و `MAE` انجام داده و نمودار این معیارها را به ازای تعداد ۳۰ و ۵۰ و ۱۰۰ اپیاک رسم نمایید. (۶ نمودار)

ج) به ازای داده های `test` مقدار هزینه اقامت در هتل `ADR` را پیش بینی نمایید و در یک فایل `CSV` جدید به نام `Results.csv` مقدار واقعی داده `ADR` و مقدار تخمین زده شده `ADR` و اختلاف بین مقدار مقدار واقعی و مقدار تخمین زده شده را بدست آورید. (۳ ستون)

د) در این سوال اطلاعات `feature` های مختلفی (مانند `Meal`, `IsCanceled`, ...) از یک اقامت مشتری هتل داریم. جمع آوری این اطلاعات هزینه بر است و هدف این هست که تنها داده های مهم هر اقامت را نگهداری کنیم که تا حد خوبی با استفاده از آنها بتوان به پیش بینی های دقیق تر هزینه اقامت واقعی بدون افزایش پیچیدگی شبکه دست یافت. چطور میتوان کشف کرد که کدام `feature` ها (ستون) در داده های ما مهم تر است؟ دقت

نمایید برای این سوال باید یک پاسخ کلی ارائه نمایید و راه حل ارائه شده نباید بر اساس پیش فرض های ذهنی از داده های آن مسئله باشد.

ه) را حل پیشنهادی خود در قسمت (د) را پیاده سازی نمایید و با استفاده از آن اهمیت **feature** های مهم را بدست آورده و آن ها را بیان نمایید. بار دیگر به ازای ۳۰ اپیاک شبکه خود را پیاده سازی کرده و مانند قسمت ج نتایج را در فایل **SelectedResult.csv** ذخیره نمایید.

سوال دوم: هدف از این تمرین ایجاد یک طبقه بند برای طبقه بندی مجموعه داده CIFAR-10 با استفاده از شبکه های MLP است. این مجموعه داده شامل ۶۰ هزار تصویر رنگی است که در ۱۰ کلاس دسته بندی شده است.

به طور معمول از ۵۰ هزار تصویر آن به عنوان مجموعه داده آموزشی استفاده میکنند و ۱۰ هزار تصویر را به عنوان مجموعه داده تست استفاده می کنند. شما این مجموعه را دانلود و ۱۰ تصویر ابتدای آن را همراه با نام آن شی نمایش داده و سپس داده ها را به ۳ بخش آموزش، تست و ارزیابی تقسیم کنید.

با استفاده از کتابخانه Keras نیز میتواند مجموعه داده را دانلود نمایید.

```
from keras.datasets import cifar10
(x_train, y_train), (x_test, y_test) = cifar10.load_data()
```

در ادامه پیش پردازش های لازم را انجام دهید تا داده ها برای آموزش شبکه عصبی آماده بشوند.

حال با توجه به آموخته های کلاس درس ، یک شبکه MLP طراحی نمائید. هدف این سوال پیاده سازی شبکه عصبی ، بررسی تاثیر تغییرات هایپرپارامترها و حل چالش مجموعه داده نامتوازن است.

فرضیات مسئله:

- تعداد لایه های مخفی را برابر ۲ در نظر بگیرید.
 - از روش Stochastic mini batch based استفاده نمایید.
- موارد زیر را در گزارش برای قسمت ب و ج بیاورید :
- در دو نمودار جداگانه تغییرات دقت و خطای مدل در هر دور را برای داده ی ارزیابی و آموزش حالت های خواسته شده نشان دهید.
 - همچنین خطا، دقت و ماتریس آشفتگی را برای داده ی تست محاسبه کنید.

الف) لازم است برای حل این مسئله از روش Stochastic mini batch based استفاده شود ، از سه دسته با اندازه های ۳۲ و ۶۴ و ۲۵۶ استفاده نمایید و تاثیر تفاوت اندازه دسته ها را در دقت و زمان آموزش شبکه بررسی نمایید.

ب) توابع فعالساز هرلایه را تغییر دهید و تاثیر توابع فعالساز را در دقت آموزش شبکه بررسی نمایید، مجموعاً ۳ مرتبه توابع فعالساز را در لایه های ماقبل آخر تغییر دهید و نتایج آن را در گزارش بیاورید.

مزایا و معایب این توابع فعالساز را نسبت به دیگری بررسی نمایید.

نکته : از توابع فعالساز ReLU ، TanH و Sigmoid استفاده کنید.

نکته : دقت نمایید در این آزمایش ها بقیه هایپرپارامترها ثابت هستند و از بهترین مدل قسمت الف استفاده نمایید.

ج) تابع خطا شبکه را تغییر دهید و تاثیر تابع خطاهای متفاوت را در دقت آموزش شبکه بررسی نمایید، مجموعاً ۲ مرتبه تابع خطا را تغییر دهید و نتایج آن را در گزارش بیاورید. دلیل این تفاوت را از منظر ریاضی بررسی نمایید.

نکته : تابع خطا را از دو خانواده متفاوت انتخاب کنید ، به طور مثال تابع خطای Cross entropy را با تابع خطای MSE میتوانید مقایسه کنید.

موعد تحویل: چهارشنبه ۱۰ دی ماه

روش تحویل: آپلود در **courses**

گزارش: برنامه های نوشته شده در MATLAB یا پایتون به همراه نتایج و شکل های خواسته شده، تحلیل نتایج و نمودارها را ضمیمه کنید.

توجه: استفاده از کتابخانه ها در حل این سوال مجاز است.

توجه: کل تمرین و گزارش باید بصورت انفرادی نوشته شود، و به برنامه ها و گزارش های کپی شده نمرهای تعلق نمیگیرد.

پیروز و سلامت باشید.