

پرسش ۴ – Multi-Layer Perceptron

در این سوال یک دیتاست برای پیشبینی قیمت در اختیار شما گذاشته شده است. ابتدا با دیتا ها کار کرده و با Feature Engineering آشنا خواهید شد و سپس با استفاده از چند MLP قیمت را پیشبینی کرده و مقایسه خواهید کرد.

هدف از این سوال آشنایی با MLP و کتابخانه های TensorFlow/Keras یا PyTorch میباشد.

۴-۱. آشنایی و کار با دیتاست (پیش پردازش)

(۴۰ نمره)

هدف از این قسمت آشنا شدن با فراخوانی داده ها و پیش پردازش آنها میباشد. برای این کار باید به ترتیب کار های زیر را انجام دهید:

- I. فایل csv مربوط به سوال را خوانده و سپس تابع info. را از Pandas فراخوانی کنید.
- II. تعداد داده هایی که Nan هستند را برحسب هر ستون نمایش دهید.
- III. از ستون CarName نام شرکت را جدا کرده و به نام CompanyName ذخیره کنید. سپس ستون های CarName، car_ID و symboling را حذف کنید. تعدادی از نام های شرکت ها اشتباه تایپ شده است. آنها را پیدا کرده و اصلاح کنید.
- IV. داده های توصیفی را به داده های عددی تبدیل کنید (راهنمایی: مثلا نوع سوخت میتواند گازی یا دیزلی باشد و میتواند ۰ و ۱ شود. میتوانید برای اینکار از pd.get_dummies استفاده کنید).
- V. Correlation Matrix را رسم کنید. چه فیچری با قیمت Correlation بیشتری دارد؟
- VI. نمودار توزیع قیمت و نمودار قیمت برحسب فیچری که بیشترین Correlation با قیمت دارد را رسم کنید.
- VII. داده ها را به train/test تقسیم کنید. (15% : 85%)
- VIII. به کمک MinMaxScaler داده های train و test را scale کنید. (توجه داشته باشید از داده های test نباید برای scale کردن داده استفاده کنید چون در این صورت leakage data رخ می دهد).

۴-۲. Multi-Layer Perceptron

(۶۰ نمره)

در این قسمت می‌خواهیم تاثیر تغییر برخی پارامترها روی شبکه را ببینیم. در این حالت نیز به ترتیب کارهای زیر را انجام دهید:

- I. ابتدا ۳ مدل MLP ساده به ترتیب با ۱، ۲ و ۳ لایه پنهان بسازید.
- II. دو Optimizer و Loss Function مختلف را بررسی کنید.
- III. با یکی از آنها هر ۳ مدل را آموزش داده و نمودار loss و loss validation را رسم کنید. معیار R^2 score را توضیح داده و سپس به کمک این معیار بهترین شبکه را انتخاب کنید.
- IV. سپس برای بهترین شبکه سه حالت باقی‌مانده (دو Optimizer و Loss Function) را بررسی کنید.
- V. پنج داده را به تصادفی از test set انتخاب کرده و قیمت را پیشبینی کنید. سپس مشخص کنید چقدر این پیشبینی شما با قیمت واقعی تفاوت دارد.