

پیش بینی قیمت خودرو از روی سن خودرو

فایل car_price.xlsx در اختیار شما قرار گرفته.

این فایل حاوی ۱۰۰۰ قلم، داده های مربوط به قیمت یک خودروی فرضی در محدوده سنی ۱ تا ۳۵ سال می باشد .

الف) داده های مورد نظر را از فایل اکسل خوانده و آنها را به نسبت ۷۵٪ و ۲۵٪ برای داده های آموزشی و تست تقسیم کنید سپس داده های آموزشی و تست را با دو رنگ متفاوت بر روی نمودار نمایش دهید.

ب) به کمک رگرسیون خطی و با استفاده از داده های آموزشی مدلی برای پیش بینی قیمت خودرو از روی سن آن ، بسازید و سپس مقدار تابع هزینه ی مدل را برای داده های آموزشی و تست محاسبه و نمایش داده و خط رگرسیون را به نمودار بخش الف اضافه کنید.

ج) به کمک رگرسیون چند جمله ای درجه ۲ ، ۳ و ۴ نیز مدل مورد نظر را ساخته و مقادیر تابع هزینه این مدل ها را نیز به صورت جداگانه برای داده های آموزشی و تست به دست آورید و ترسیم مدل را به نمودار بخش الف اضافه کنید.

د) با توجه به مقادیر به دست آمده برای توابع هزینه و نمودارهای رسم شده در قسمت ب و ج ، به نظر شما کدام مدل مناسب تر است.

ه) با انجام این پروژه و دیدن نتایج ، نظر خود را در مورد پیچیده کردن مدل بگویید.

و) اگر دادهای مسئله بسیار کم باشند (مثلاً به جای ۱۰۰۰ قلم، ۱۰ قلم) ، آنگاه با پیچیده کردن مدل چه اتفاقی رخ می دهد؟

توجه :

1- برای بخش ایجاد و آموزش دادن مدل ها از کتابخانه های آماده نظیر scikit-learn استفاده فرمایید.

2- نیازی به نمایش همه مدل ها بر روی یک نمودار نیست و می توانید پس از ایجاد و نمایش هر مدل نتایج آن را ثبت کرده یا به خاطر بسپارید و با تغییر درجه چند جمله ای، برنامه را برای دیدن نتایج بعدی مجدداً اجرا کنید.

موفق باشید (:)