

تکلیف سری ۱:

زمان تحویل: به عهده خواننده 😊

در مسئله زیر قصد داریم از کتابخانه `scikit` برای تخمین قیمت خودرو تویوتا استفاده کنیم. دیتاست مذکور تحت عنوان یک فایل با نام `ToyotaCorolla.csv` در اختیار شما قرار میگیرد.

ابتدا ۷۰ درصد داده‌گان را برای آموزش و مابقی را برای تست قرار دهید. (میتوانید برای تخمین دقیق تر از خطا از `cross validation` استفاده کنید) سپس از مدل های رگرسیون برای تخمین قیمت استفاده کنید.

برای بهبود دقت در صورت لزوم میتوانید از مدل های غیرخطی استفاده کنید و همچنین استفاده از روش های `feature selection` موجود در `scikit` و `statsmodels` نمره مثبت خواهد داشت.

در صورت علاقه خیلییییییی زیاد میتوانید رگرسیون خطی را تنها با استفاده از `numpy` و `pandas` به روش گرادینان کاهشی پیاده کنید و برای همگرایی زودتر این روش `stochastic gradient descent` کاراتر است.

در صورت علاقه خیلی خیلییییییی بیشتر رگرسیون `ridge` و `lasso` را برای این دیتاست انجام دهید و تغییرات میزان خطا با تغییر ضریب رگولاریزاسیون بررسی کنید .

توجه: برای در نظر گرفتن معیار خطا `mse` لزوما مناسب نیست ولی برای شروع خوب است. میدانید چرا؟