



یادگیری ماشین

تابستان ۱۴۰۳

مدرس: دکتر مهدی جعفری

Linear Regression

تمرین سری اول (۱۰۰ نمره)

زمان تحویل: ۱۴ شهریور

سوالات عملی (۱۰۰ نمره)

سوال ۱: (۱۰ نمره)

هدف این مساله آموزش یک مدل رگرسیون برای تخمین قیمت ماشین‌های موجود در مجموعه داده‌ی Toyota Corolla بر حسب ویژگی‌های مختلف آن است. سه بخش، `test`، `validation` و `data` در سه فایل جداگانه در پوشه `data` در اختیار شما قرار داده شده است. همچنین در فایل `data/data codes.xlsx` توضیحاتی در باره‌ی هریک از ویژگی‌ها نوشته شده است.

(آ) در مرحله‌ی اول یک مدل Linear Regression را با استفاده از کتابخانه `scikit-learn` آموزش داده و دقت خود را برای `learning rate = {1, 0.5, 0.1, 0.01, 0.001}` بر روی مجموعه داده‌های `validation` گزارش مقدار و بهترین مقدار را انتخاب کنید

(ب) در مرحله‌ی دوم مدل Linear Regression را بدون استفاده از کتابخانه `scikit-learn` آموزش داده (یعنی مدل را از پایه پیاده سازی کنید) و دقت خود را برای `learning rate = {1, 0.2, 0.2, 0.02, 0.002}` بر روی مجموعه داده‌های `validation` گزارش داده و مجدداً بهترین مقدار را انتخاب کنید

(ج) در مرحله آخر با استفاده از مدلی که از پایه نوشته شده (از مرحله پیشین) و تغییر در داده‌ها یک مدل خطی درجه ۳ آموزش دهید و دقت نهایی را بر روی قسمت `test` گزارش دهید.

توضیحات:

- ابتدا همه‌ی ویژگی‌های `categorical` را به `one-hot` تبدیل کنید.
- برای قسمت ب فقط می‌توانید از پکیج‌های `numpy` و `pandas` استفاده کنید.
- برای رسم نمودارها لازم است اطلاعات مورد نیاز مانند `title`، `axis label` و `legend` را تنظیم کنید.
- فایل نهایی را در قالب یک فایل `Jupyter Notebook` تحویل دهید و همچنین توضیحات لازم را در داخل نوتبوک گزارش دهید.