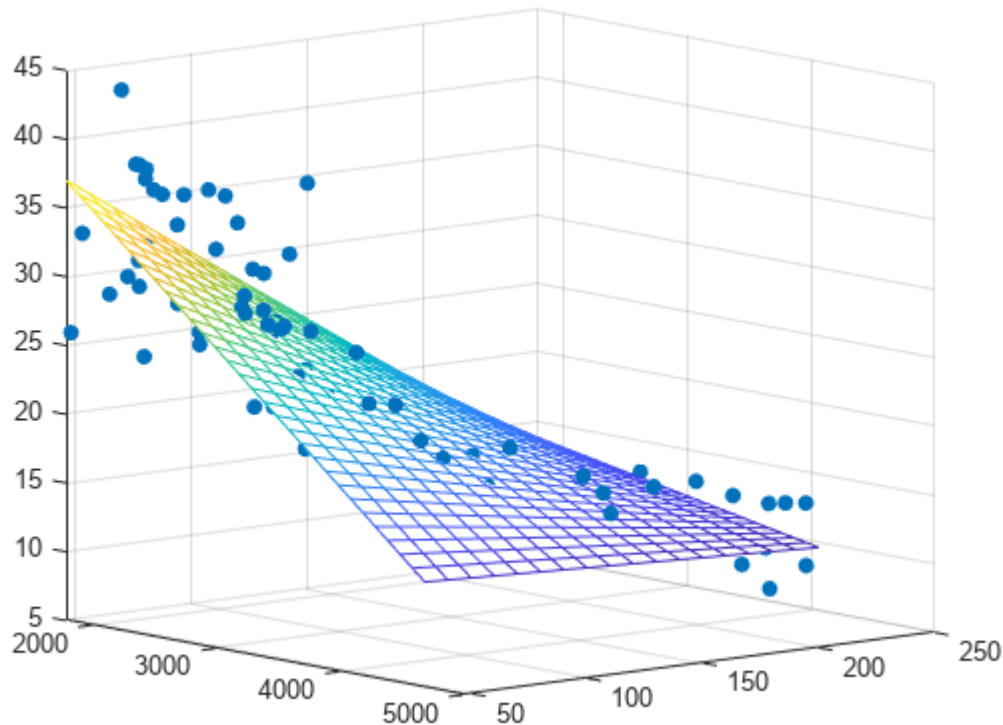


رگرسیون خطی چند متغیره



صورت سؤال:

در این سؤال، هدف برازش (*fit*) یک رگرسیون خطی چند متغیره بر روی مجموعه‌ای از نقطه‌های (x_1, \dots, x_6, y) داده شده در ورودی است به نحوی که بیش‌ترین تطابق با مجموعه نقطه‌ها حاصل شود. رابطه کلی یک رگرسیون خطی چند متغیره را می‌توانیم به‌صورت زیر در نظر بگیریم.

$$f(x_0, x_1, \dots, x_n) = w_0 \cdot x_0 + w_1 \cdot x_1 + \dots + w_n \cdot x_n + b$$

در حل این مسئله، هدف پیدا کردن مقادیر مناسب برای پارامترهای w_1 تا w_6 و b است. به این منظور باید الگوریتم نزول در راستای گرادیان (*gradient descent*) توسط شما پیاده‌سازی گردد.

▼ تذکر مهم!

در حل این تمرین نباید از پیاده‌سازی‌های آماده استفاده شود. ضروریست این الگوریتم‌ها توسط شما پیاده شوند.

ورودی:

ورودی شامل یک دیتاست به فرمت فایل CSV است که در هر سطر آن قیمت یک پرواز با توجه به ویژگی‌های آن ذکر شده است.

- داده‌های موردنظر از طریق این آدرس قابل دانلود است.

خروجی:

خروجی برنامه شما باید یک فایل به فرمت txt باشد که حاوی موارد زیر است:

- پارامترهای پیدا شده برای ترکیب خطی ویژگی‌ها
- زمان مورد نیاز برای آموزش مدل بر حسب ثانیه
- گزارش میزان خطا بر روی مجموعه تست برحسب معیارهای زیر:

MSE ◦

RMSE ◦

MAE ◦

R2 ◦

خروجی نمونه

PRICE = -15 * [Feature_Name_0] + 2 * [Feature_Name_1] + 3 * [Feature_Name_2] + ...

Training Time: 4s

Logs:

MSE: 2322

RMSE: 123

MAE: 11

R2: 0.67

نام فایل خروجی لازم است که به فرمت زیر باشد:

[GroupNumber]-UIAI4021-PR1-Q2.txt

▼ راهنمایی

- دقت کنید متغیرهای دسته‌ای برای انجام آموزش نیاز دارند به مقدار عددی تبدیل شوند. برای متغیرهای ترتیبی می‌توانید به هر رشته یک عدد صحیح با توجه به اولویت آن اختصاص دهید و برای متغیرهای اسمی می‌توانید از روش *one-hot encoding* استفاده کنید.
- با کمک تابع `train_test_split` در کتابخانه `sklearn` ، ۲۰ درصد از مجموعه داده را به عنوان داده تست برای ارزیابی مدل جدا کنید.