

○ داده‌های پروژه:

<http://archive.ics.uci.edu/ml/datasets/Auto+MPG>

○ توضیحات پروژه:

داده‌های مربوط به میزان مصرف سوخت اتومبیل‌های سه ناحیه آمریکا، ژاپن و اروپا در دست است که رکورد هر ناحیه توسط اعداد ۱ تا ۳ در ستون *Origin* برچسب گذاری شده‌اند. هدف پیش‌بینی مصرف سوخت اتومبیل با استفاده از مدل رگرسیون خطی است.

مدل کلی رگرسیون خطی به صورت زیر است:

$$y = w_0 + \sum_{i=1}^m w_i \phi_i(x_i)$$

با توجه به اینکه مسئله دارای ۷ متغیر ورودی است، بسط نیوتن حاصل از این ۷ متغیر دارای پارامترهای زیادی خواهد بود که برای ساده‌سازی از جملات ضربی متغیرها صرف نظر شود. بنابراین مدل ساده شده رگرسیون فوق به صورت زیر خواهد بود:

$$y = w_0 + \sum_{i=1}^m w_{1i}x_1^i + w_{2i}x_2^i + w_{3i}x_3^i + w_{4i}x_4^i + w_{5i}x_5^i + w_{6i}x_6^i + w_{7i}x_7^i$$

○ خواسته‌ها:

۱. درجه بهینه کرنل چندجمله را به روش Elbow محاسبه کنید.

(راهنمایی: خطای تست و خطای آموزش را نسبت به درجات مختلف چندجمله‌ای رسم کنید)

۲. تعداد بهینه توابع کرنل RBF را به روش Elbow تعیین کنید.

(راهنمایی: خطای تست و خطای آموزش را نسبت به تعداد توابع کرنل رسم کنید)

۳.

۱. اگر درجه چندجمله‌ای را برابر با ۲۵ در نظر بگیریم، مقدار بهینه ضریب تنظیم‌کنندگی را تعیین کنید.
۲. اگر تعداد توابع RBF را برابر با ۹۰ در نظر بگیریم، مقدار بهینه ضریب تنظیم‌کنندگی را تعیین کنید.

۴. با استفاده از روش LASSO و استفاده از توابع آماده پایتون، تاثیر هر متغیر ورودی را بر پیش‌بینی مصرف سوخت (برای هر دو حالت کرنل چندجمله و کرنل RBF) تعیین کنید.

(درجه چندجمله را برابر با ۱۰۰ و تعداد توابع RBF را نیز برابر با ۱۰۰ در نظر بگیرید).