

به نام خالق عقل و خرد



جلسه اول

یادگیری ماشین

نیم سال دوم ۱۴۰۴ - ۱۴۰۳

دانشکده ریاضی و آمار

دانشگاه اصفهان

اسفند ماه ۱۴۰۳

استاد درس:

دکتر فاطمه منصوری

عنوان تمرین : پیاده‌سازی Regression

هدف: درک و پیاده‌سازی تکنیک‌های مختلف Regression ، انجام پیش‌پردازش داده‌ها، ارزیابی عملکرد مدل با استفاده از معیارهای مناسب و تفسیر و تحلیل خروجی‌ها

دستورالعمل :

۱. دیتاست پیش‌بینی قیمت خودرو (car.csv) را از گیت‌هاب درس دانلود کنید.
۲. مجموعه داده را در یک Dataframe پانداس بارگزاری کرده و چند داده اول آنرا مشاهده نمایید.
۳. اطلاعات کل دیتاست مانند ویژگی‌ها، نوع آنها، مقادیر گمشده و ... را مشاهده نمایید.
۴. ارتباط بین ویژگی‌ها و قیمت را با استفاده از scatter plot و جدول correlation مشاهده کنید.
۵. پیش‌پردازش داده‌ها مانند حذف مقادیر گم شده یا حذف مقادیر categorical و نرمال کردن مقادیر ویژگی‌های عددی را انجام دهید. (برای نرمال کردن مقادیر یک تابع بنویسید که نرمال‌سازی به روش Min-Max انجام دهد).
۶. داده‌ها را به دو دسته آموزش و تست تقسیم کنید
۷. الگوریتم linear regression را پیاده‌سازی نمایید. و برای هر ویژگی مقدار وزن اختصاص داده شده به آنرا به دست آورید.
۸. ارزیابی مدل
توابع زیر را پیاده‌سازی کنید و با استفاده از آنها مدل خود را ارزیابی کنید.
Mean Absolute Error □
Mean Squared Error □
۹. با استفاده از K-fold cross validation مدل خود را ارزیابی کنید.
۱۰. برای ویژگی‌ها polynomial feature ایجاد کنید و یک مدل polynomial regression به داده‌ها اختصاص دهید.
۱۱. مدل‌های مختلف چند جمله‌ای را امتحان کرده و مدلی که نسبت به دیگر مدل‌ها مناسب‌تر است گزارش کنید. مدل بهتر را با توجه به بررسی overfitting, underfitting انتخاب کنید.
۱۲. گزارشی از نتایج و مقایسه‌های خود برای هر قسمت تهیه و ارائه نمایید.