

۱. مقدمه

هدف پیاده سازی الگوریتم های linear regression به صورت دستی است.

۲. تسک های خواسته شده

پیش تر دیتا را پاکسازی کرده بودیم و مقادیر null را پر و داده های outlier را حذف کرده بودیم. سه فیچری که در صورت تمرین خواسته شده را جدا کرده و دیتاست جدیدی ایجاد می کنیم. از بین این داده ها، یکی از آنها کتگوریکال است که با استفاده از one-hot عددی می کنیم. حال داده ها را به دو بخش آموزشی و آزمایشی تقسیم کرده و داده ها را scale می کنیم. حال linear regression را به صورت دستی پیاده سازی می کنیم. دقت مدل برابر ۲۶,۲۲٪ می شود. حال با پیاده سازی پکیج linear regression به دقت ۴۰,۰۷٪ می رسم. با پیاده سازی پکیج lasso regression به دقت ۴۰,۰۵٪ می رسم و با پیاده سازی پکیج ridge regression به دقت ۴۰,۰۷٪ می رسم.

در رگرسیون ridge و lasso به علت وجود ترم بهینه کننده، وزن های مدل از یک حدی بالا تر نمی روند و بنابراین مدل به خوبی fit نمی شود. در ضمن اگر رابطه به صورت خطی نباشد، بایاس زیاد است و در نتیجه مدل underfit میشود. اما به دلیل وجود همین ترم ها، وزن خیلی از فیچر ها نزدیک به صفر می شود که می تواند تفسیر خوبی از روند دیتا به ما بدهد. به صورت کلی مانند شکل هر چقدر انعطاف پذیری مدل افزایش یابد، تفسیر پذیری آن کاهش می یابد و بالعکس.

