

تسک موسسه گزینه دو بررسی دو روش رگرسیون خطی

۱ مقدمه

Lasso Regression یک الگوریتم یادگیری ماشین است که برای انجام Regression (یا پیش‌بینی یک متغیر پیوسته) استفاده می‌شود. یکی از ویژگی‌های اصلی Lasso Regression این است که از نرم یک (L1 norm) برای اعمال تابع هزینه (cost function) استفاده می‌کند.

Ridge Regression یک الگوریتم یادگیری ماشین است که برای انجام Regression (یا پیش‌بینی یک متغیر پیوسته) استفاده می‌شود. یکی از ویژگی‌های اصلی Ridge Regression این است که از نرم دو (L2 norm) برای اعمال تابع هزینه (cost function) استفاده می‌کند.

۲ توضیحات

هدف از استفاده از نرم یک در تابع هزینه Lasso Regression این است که موجب کاهش مقادیر برخی از پارامترها به صفر می‌شود، که این به تنهایی باعث انجام یک فرآیند انتخاب ویژگی (feature selection) می‌شود. از آنجایی که نرم یک مقادیر پارامترهای مدل را مستقیماً کنترل می‌کند و این مقادیر را به صفر می‌برد، می‌توانیم ببینیم که چگونه این روند می‌تواند به تنظیم MAE (Mean Absolute Error) کمک کند. MAE نیز تابعی از فاصله میان مقادیر پیش‌بینی شده توسط مدل و مقادیر واقعی برچسب‌ها است. با کاهش تعداد ویژگی‌هایی که به صورت مستقیم در مدل استفاده می‌شوند (با صفر کردن برخی از پارامترها)، ممکن است تغییرات مدل کمتری رخ دهد و در نتیجه MAE نیز به صورت کلی کاهش یابد. این انتخاب ویژگی‌ها و تنظیم MAE را در کنار هم می‌توان به عنوان یک فرآیند بهینه‌سازی در نظر گرفت. با این حال، لازم به ذکر است که هرچقدر مقدار α بزرگتر باشد، تاثیر نرم یک بر روی اندازه پارامترها و بر این تغییرات کمتر خواهد بود. این توازن بین تنظیم ویژگی‌ها و کاهش مقدار MAE را تعیین می‌کند.

$$J(0) = \frac{1}{2m} \sum_{j=1}^n (h_0(x^{(i)}) - y^{(i)})^2 + \frac{a}{2} \sum_{j=1}^n |\theta_j| \quad (1)$$

هدف از استفاده از نرم دو در تابع هزینه Ridge Regression این است که موجب کاهش مقادیر پارامترها شود و این باعث کنترل ویژگی‌های مختلف مدل می‌شود. از طریق اعمال نرم دو، مقادیر پارامترها به شدت تنظیم می‌شوند و این باعث می‌شود که مقدار RMSE (Root Mean Squared Error) نیز به صورت کلی کمتر شود. با این حال، لازم به ذکر است که هرچقدر مقدار α بزرگتر باشد، تاثیر نرم دو بر روی اندازه پارامترها و بر این تغییرات کمتر خواهد بود. این توازن بین تنظیم ویژگی‌ها و کاهش مقادیر RMSE را تعیین می‌کند.

$$J(0) = \frac{1}{2m} \sum_{j=1}^n (h_0(x^{(i)}) - y^{(i)})^2 + \frac{a}{2} \sum_{j=1}^n \theta_j^2 \quad (2)$$

۳ نتیجه گیری

از آن جایی که در فرمول ریاضی Lasso Regression عبارت دوم معادله یک عبارت نرم یک و مشتق ناپذیر بوده بنابراین حد فاصل داده ها در MAE با نرم یک کمتر بوده در نتیجه مقدار MAE کمتر خواهد بود. همین اتفاق به طور برعکس برای Ridge Regression به علت مشتق پذیر بودن عبارت دوم مقدار RMSE بهتر خواهد بود بنابراین، برای کمترین MAE ممکن است Lasso Regression مناسب باشد، زیرا این روش تمایل به حذف ویژگی‌های غیرضروری و تنظیم وزن‌ها به صفر دارد، در حالی که برای کمترین RMSE ممکن است Ridge Regression مناسب باشد، زیرا این روش می‌تواند با کاهش واریانس وزن‌ها، دقت بالاتری در پیش‌بینی‌ها فراهم کند.