

تمرین دوم مبانی یادگیری ماشین (بخش 2)

محمد رضا ضیا لاری (97222057)

- 1- رگرسیون Ridge : در رگرسیون ridge جمله ای که برای regularization اضافه میشود همان نرم 2 ضرایب ویژگی های ما می باشد .
رگرسیون Lasso : در این رگرسیون که شباهت زیادی به رگرسیون Ridge دارد به جای نرم 2 ، نرم 1 ضرایب ویژگی به عنوان پناستی اضافه می شود .
از هر دو این رگرسیون ها برای کم کردن خطا استفاده می کنند اما تفاوت اصلی بین این دو که از تفاوت در نرم 1 و نرم 2 بودن حاصل می شود در این است که در رگرسیون لاسو ممکن است ویژگی هایی که تاثیر کمتر دارند و یا حتی بی تاثیر هستند ضرایب 0 بگیرند اما در رگرسیون ریج این مقدار به صفر میل می کند و هیچ گاه برابر 0 نمی شود.(منبع فرادرس)
- 2- یکی از راه ها برای پیدا کردن آن، پیاده سازی مدل توسط چندین آلفا متفاوت به صورت cross validation است. به این صورت که توسط لیستی از آلفا های مختلف مدل را ارزیابی میکند و بهترین نتیجه را میدهد .
- 3- باافزایش فولد بایاس کاهش می یابد زیرا فولدها به مجموعه کل داده نزدیکتر خواهند شد اما هر چه به Leave-One-Out-Cross-Validation نزدیک تر می شویم واریانس و مدت زمان اجرا افزایش می یابد .
- 4- نوع خاصی از cross validation است که در آن به دلیل عدم وجود دیتای تست یا دلایل دیگر ریا، بخشی از دیتای ترین را در هر مرحله کنار می گذاریم و از آن به عنوان دیتای تست و سنجش خطا و دقت استفاده می کنیم .

5- این روش برای نمونه گیری از داده اصلی است که با جایگذاری انجام میشود یعنی هر داده ای که وارد نمونه گیری میشود دوباره پس از آنکه کار آن تمام شد به دیتاست برمیگردد و باز هم میتواند در نمونه گیری های بعدی شرکت داده شود و سپس میانگین دقت و خطا را برای سنجش در نظر می گیریم . تفاوت این روش با cross validation این است که در cross validation ما بخشی از دیتا که برای سنجش است را به کل از مجموعه داده جدا می کنیم اما در BootStrapping اینگونه نیست.

6- در 2-fold cross-validation داده ها را بصورت رندوم به 2بخش هم اندازه تقسیم می کنیم. یک بخش به عنوان داده ی تست و دیگری بعنوان داده ی ترین استفاده می شود و بار دیگر همین کار را با بخش دیگر انجام مید هیم در نتیجه دو معیار خطای و دقت خواهیم داشت که می توانیم از آن ها میانگین بگیریم و بعنوان خطای و دقت مدل گزارش دهیم. حال همین کار را 5 بار مکررا انجام می دهیم و برای سنجش میانگین نتایج خطا و دقت این 5 مرتبه را گزارش می دهیم . این کار احتمالا باعث می شود که خطا و دقت بهتر و دقیق تر یرا گزارش دهیم .