

۱- در رگرسیون Ridge و Lasso ما به دنبال پرهیز از overfit شدن مدل‌مان روی دیتا‌ها هستیم برای همین در لاس فانکشن این رگرسیون‌ها یک ترم پنالتی اضافه می‌کنیم.

در رگرشن Ridge ما مجموع مربعات وزن‌ها که در یک هاپر پارامتر α ضرب می‌شود را به عنوان پنالتی به لاس فانکشن اضافه می‌کنیم.

در رگرشن Lasso ما مجموع قدر مطلق وزن‌ها ضرب در α را به عنوان پنالتی در نظر می‌گیریم.

از رگرسیون Ridge بیشتر برای دیتا‌های کوچک استفاده می‌شود چون مربعات وزن‌ها نباید خیلی زیاد شود.

از رگرسیون Lasso هم بیشتر برای دیتا‌هایی که وزن sparse یعنی تعداد زیادی صفر داشته باشد استفاده می‌شود.

۲- می‌توانیم با آزمون و خطا α ای مناسب را برای ضریب رگولاریزیشن پیدا کنیم اما می‌توانیم از روش‌هایی مثل cross validation استفاده کنیم و در حین لرن شدن دیتا‌هایمان α هم‌ترین شود.

۳- افزایش فولد‌ها باعث می‌شود ما در هر مرحله‌ی ترین کردن دیتای‌ترین بیشتری داشته باشیم و همچنین امکان غیر واقعی بودن دقت را هم به حداقل می‌رسانیم اما هم هزینه‌ی زمانی ما را بالا می‌برد و هم دیتای تست ما نباید از به حدی کمتر شود وگرنه ولیدیشن درستی نخواهیم داشت

۴- این روش به این شکل است که ما اگر n دیتا داریم آن‌را به n فولد تقسیم کنیم.

این روش روی دیتا‌هایی که مقدار کمی از آنها در دسترس است به کار می‌رود یا دیتا‌هایی که دقت در آن‌ها خیلی مهم تر از هزینه‌ی زمانی است چون اگر دیتا زیاد باشد این روش هزینه‌ی زمانی بسیار زیادی می‌برد.

۵- بوتسترپینگ یک روش انتخاب نمونه است که در آن همانند کراس ولیدیشن عمل می‌کنیم با این تفاوت که در کراس ولیدیشن فولد‌های ما مرز‌های مشخصی داشتند و اشتراکشان صفر بود اما در بوتسترپینگ به صورت رندوم نمونه‌ها را مشخص می‌کنیم.