



درس : یادگیری ماشین

دانشجو : امیرمحمد خرازی

شماره دانشجویی : ۴۰۱۵۲۵۲۱۰۰۲

استاد درس : دکتر منصور رزقی آهق

دانشکده علوم ریاضی ، گروه علوم کامپیوتر، گرایش داده‌کاوی

پرسش‌های کلاسی سری چهارم

گیت‌هاب این پرسش ( لینک )

گیت‌هاب درس ( لینک )

## گزارش و توضیحات :

توضیحات کامل داخل کدها آورده شده است و در اینجا نیازی به توضیحات بیشتر نیست.  
در جایی از سوال که از ما خواسته است منظم سازی با استفاده نرم ۲ را به کمک روش‌های ماتریسی حل کنیم داریم :  
مسئله :

$$\min ||\Phi W - t||_2^2 + \lambda ||W||_2^2$$

معادل است با :

$$\min ||[\Phi, \sqrt{\lambda}I]^T W - [t, 0]^T||_2^2$$

من برای اینکه به صورت ماتریس ستونی ننویسم همه را یک transpose گرفتم.  
دلیل ؟

می‌دانیم  $||X||_2^2 = X^T X$  فرض کنید معادله بالا بصورت زیر است  $||A||_2^2$  که در آن  $A = [\Phi, \sqrt{\lambda}I]^T W - [t, 0]^T$   
حال جواب آن برابر خواهد بود با  $A^T A$  که یعنی  $([\Phi, \sqrt{\lambda}I]^T W - [t, 0]^T)^T ([\Phi, \sqrt{\lambda}I]^T W - [t, 0]^T)$

$$\Rightarrow ([\Phi, \sqrt{\lambda}I]^T W - [t, 0]^T)^T ([\Phi, \sqrt{\lambda}I]^T W - [t, 0]^T)$$

$$\Rightarrow ([\Phi W - t, \sqrt{\lambda}W]^T)^T ([\Phi W - t, \sqrt{\lambda}W]^T)$$

$$\Rightarrow ([\Phi W - t, \sqrt{\lambda}W]) ([\Phi W - t, \sqrt{\lambda}W]^T)$$

$$\Rightarrow ([\Phi W - t][\Phi W - t]^T) + \sqrt{\lambda}W\sqrt{\lambda}W^T$$

$\Rightarrow$  هر جا ترانواده است، عملاً نیست و هر جا نیست، عملاً هست.

$\Rightarrow$  به این دلیل اینطوری شد که از اول همه را بصورت ترانواده گرفته بودم.

$\Leftrightarrow$  بخش اول تعریف نرم ۲ و رگرسیون است و بخش دوم تعریف نرم ۲ و منظم ساز است

$$\Rightarrow ||\Phi W - t||_2^2 + \sqrt{\lambda}\sqrt{\lambda}||W||_2^2 = ||\Phi W - t||_2^2 + \lambda||W||_2^2$$

لذا معادل هستند.

جواب این بهینه سازی بصورت  $(\Phi^T \Phi + \lambda I)W = \Phi^T t$  است. دلیل ؟ فرض کنید تابع هدف بصورت زیر است :  $\|AW - T\|_2^2$  که در آن  $A = [\Phi, \sqrt{\lambda}I]^T$  و  $T = [t, 0]^T$  . جواب این بهینه سازی، با حل دستگاه زیر صورت می گیرد :  $(A^T A)W = A^T T$  داریم :

$$\begin{aligned} \Rightarrow (A^T A)W &= A^T T \\ \Rightarrow \left( [\Phi^T, \sqrt{\lambda}I][\Phi, \sqrt{\lambda}I]^T \right) W &= [\Phi^T, \sqrt{\lambda}I][t, 0]^T \\ \Rightarrow (\Phi^T \Phi + \lambda I) W &= (\Phi^T t + 0) \\ \Rightarrow (\Phi^T \Phi + \lambda I) W &= (\Phi^T t) \end{aligned}$$

همچنین نتیجه ای که از کدها گرفتیم این بود که با نرم ۲ خیلی به داده های پرت حساس می شویم و عملاً overfit داریم ، یا خیلی روی داده های آموزش متمرکز می شویم که اندازه  $W$  خیلی زیاد می شود و روی داده های آزمایش بد عمل می کنیم که باز هم overfit داریم. این شرایط برای زمانی بود که از رگرسیون معمولی استفاده می کردیم. اگر از رگرسیون با تابع loss نرم ۱ استفاده کنیم به نقاط پرت کمتر حساسیم و در اغلب موارد نتیجه بهتری می گیریم. با منظم سازی های نرم ۱ و نرم ۲ می توانیم اندازه  $W$  را کنترل کنیم که overfit نکنیم. با نرم ۱ در منظم سازی، دیدیم که اغلب  $w \in W$  مقدار نزدیک به صفر می گیرند ولی بعضی از آن ها مقدار قابل توجهی دارند. با این روش منظم سازی می توانیم پیچیدگی مدل را کاهش دهیم، ابعاد را کمتر کنیم و غیره . که این منظم سازی  $W$  را sparse می کند. با نرم ۲ در منظم سازی، دیدیم که اغلب  $w \in W$  مقدار نزدیک به هم دارند. یعنی عملاً بیشتر  $w$  ها در یک رنج هستند. این منظم سازی  $W$  را smooth می کند. با روش ElasticNet تلاش کردیم هر دو این موارد را منظم سازی ها را به اجرا برسانیم.

## مشکلات

از نمونه به مشکلاتی که برخوردیم :

- عدم همگرایی منظم سازی نرم ۱
- عدم همگرایی منظم سازی ElasticNet و برازش تنها  $W_0$  در آن.