درس : یادگیری ماشین

دانشجو : امیرمحمد خرازی

شماره دانشجویی : ۴۰۱۵۲۵۲۱۰۰۲

استاد درس : دکتر منصور رزقی آهق

پرسش‌های کلاسی سری پنجم

دانشکده علوم ریاضی ، گروه علوم کامپیوتر، گرایش داده‌کاوی

گیت‌هاب درس (لینک)

گیت‌هاب این پرسش (لینک)

سوال ۱ :

حزئیات لازم در فایل کد آورده شده است . روش رسم خط های جدا ساز نیز در سوال ۲ کمی اشاره شده است . برای رسم بیز از کانتور استفاده شده است. مقایسه ها نیز در همان کد قابل مشاهده است.

سوال ۲ :

ابتدا دوگان مسئله SVM را بدست می‌آوریم:

$$Primal : SVM = \max \frac{||W||_2^2}{2}$$

$$s.t$$

$$y_i \times (W^T X_i + W_0) \geq 1 \Rightarrow 1 - y_i \times (W^T X_i + W_0) \leq 0$$

$$L(W, W_0, \alpha) = f_0(W) + \sum_{i=1}^{N} \alpha_i \left[1 - y_i \times (W^T X_i + W_0)\right]$$

$$if \quad \alpha_i \geq 0 \Rightarrow \min L = g(\alpha) \leq f_0(x^\star)$$

$$L(W, W_0, \alpha) = \frac{1}{2} W^T W + \sum \alpha_i - \sum \alpha_i y_i W^T X_i - \sum \alpha_i y_i W_0$$

$$\frac{\partial L}{\partial W} = W - \sum \alpha_i y_i X_i = 0 \Rightarrow W = \sum \alpha_i y_i X_i$$

$$\frac{\partial L}{\partial W_0} = \sum \alpha_i y_i = 0$$

در فرمول لاگرانژ جایگذاری می‌کنیم

$$g(\alpha) = \sum \alpha_i - \frac{1}{2} \sum \sum \alpha_i \alpha_j y_i y_j X_i X_j$$

دوست داریم g به بهنیه مسئله primal نزدیک شود پس آن را ماکسیمم می‌کنیم.

$$dual : maxg(\alpha) \quad alpha_i \geq 0$$

با حل مسئله دوگان داریم : $W = \sum \alpha_i y_i X_i$ و در ساخت W تنها آن‌هایی اثر دارند که α_i آن‌ها مخالف صفر است.

پس :

$$W = \sum \alpha_i y_i X_i = \begin{bmatrix} 0.846 \\ 0.3852 \end{bmatrix}$$

با داشتن $W^\star$ می‌توانیم با استفاده از شرط KKT ، مقدار W_0 را بدست آوریم.
می‌گوئیم $\alpha_i \times \left[1 - y_i(W^{\star T} X_i + W_0)\right] = 0$
بنابراین اگر در حایی $\alpha_i \neq 0$ بود، آنگاه بخش دوم آن صفر است .

$$\begin{aligned}
&\alpha_1 : \left[1 - y_1 \times (W^{\star T} X_1 + W_0)\right] \\
&\quad W_0^{(1)} = 1 - W^{\star T} X_1 = -3.50108 \\
&\alpha_4 : \left[-1 - y_4 \times (W^{\star T} X_4 + W_0)\right] \\
&\quad W_0^{(4)} = 1 - W^{\star T} X_4 = -3.5001999999999995 \\
&\alpha_7 : \left[1 - y_7 \times (W^{\star T} X_7 + W_0)\right] \\
&\quad W_0^{(7)} = 1 - W^{\star T} X_7 = -3.5017999999999994 \\
&\alpha_9 : \left[-1 - y_9 \times (W^{\star T} X_9 + W_0)\right] \\
&\quad W_0^{(9)} = 1 - W^{\star T} X_9 = -3.50092 \\
&\quad \Rightarrow W_0 = \frac{1}{4} \sum W_0^{(i)} = -3.5009999999999994 \approx -3.501
\end{aligned}$$

لذا معاله این ابر صفحه خواهد بود : $W^{\star T} X + W_0^\star = 0$
که بطور خلاصه برابر است با :

$$W^{\star T} X + W_0^\star = 0 \Rightarrow W_1 X_1 + W_2 X_2 + W_0 = 0 \Rightarrow \frac{W_1}{W_2} X_1 + \frac{W_2}{W_2} X_2 + \frac{W_0}{W_2} = 0$$

$$\Rightarrow X_2 = -\frac{W_1}{W_2} X_1 - \frac{W_0}{W_2} \quad \text{معادله خط جداساز}$$

برای فاصله یک نقطه مانند X_6 از این خط می‌توان یک نقطه روی این خط انتخاب کرد و فاصله آن را با نرم ۲ بدست آورد :

$$dist(X_6, X_{i2} = -\frac{0.846}{0.3852} X_{i1} - \frac{-3.5}{0.3852})$$

$$\Rightarrow d = \frac{|W_1 X_{61} + W_2 X_{62} + W_0|}{\sqrt{W_1^2 + W_2^2}} = -1.2497429918467136 \quad \text{فاصله مثبت است} \quad 1.2497429918467136$$

درون محدوده بودن آن به مسئله اصلی بر می‌گردد. بهینه مسئله اصلی را اگر بدست آوریم یعنی $||x_1 - x_2||_2 = \frac{||W||_2}{2}$ با

استفاده از مقادیر بدست آمده از W این مقدار برابر تقریبا 0.46 است که از هر دو طرف خط میانی فاصله دارد. لذا این نقطه خارج این بازه قرار خواهد داشت. جون فاصله آن بیشتر از این مقدار است.

برای کلاس بندی نقطه $z = (3,3)^T$ کافی است آن را در معادله خط گذاشته و جواب را بررسی کنیم . اگر منفی بود، کلاس ۲ و اگر مثبت بود کلاس ۱ است.

$$z : W_1X_1 + W_2X_2 + W_0 \Rightarrow = 0.19260000000000055 > 0 \Rightarrow \in C_1 : 1$$

سوال ۳ :

هدف پیدا کردن صورت دوگان مسئله زیر است :

$$\begin{aligned}
&\min_{W,b,\epsilon_i} \frac{||W||^2}{2} + C\sum \epsilon_i \\
&\quad s.t : \\
&\quad y_i(W^TX_i + b) \geq 1 - \epsilon_i \Rightarrow (1-\epsilon_i) - y_i(W^TX_i + W_0) \leq 0 \\
&\quad \epsilon_i \geq 0 \Rightarrow -\epsilon_i \leq 0
\end{aligned}$$

ابتدا فرم لاگرانژ آن را بدست می‌آوریم:

$$L(W, W_0, \epsilon, \alpha, \beta) = \frac{||W||_2^2}{2} + C\sum \epsilon_i + \sum \alpha_i \times \left[(1-\epsilon_i) - y_i(W^TX_i + W_0)\right] + \sum \beta_i \times [-\epsilon_i]$$

مرحله بعدی مشتق گرفتن و برابر صفر قرار دادن است :

$$\frac{\partial L}{\partial W} = 0 \Rightarrow W = \sum \alpha_i y_i X_i \quad \text{مانند قبل}$$
$$\frac{\partial L}{\partial W_0} = 0 \Rightarrow \sum \alpha_i y_i = 0 \quad \text{مانند قبل}$$
$$\frac{\partial L}{\partial \epsilon_i} = 0 \Rightarrow 0 + C - \alpha_i - \beta_i = 0 \Rightarrow \beta_i = C - \alpha_i \quad \text{C را ثابت گرفته ایم}$$

از آنجایی که β_i و α_i باید مثبت باشند در مسئله دوگان، داریم $C - \alpha_i \geq 0$ پس $0 \leq \alpha_i \leq C$

با بدست آوردن این مقادیر آن‌ها را در لاگرانژ جایگذاری می‌کنیم (مانند سوال ۲) تا به $g(\alpha, \beta)$ برسیم.

$$\Rightarrow \frac{W^TW}{2} + C\sum \epsilon_i + \sum \alpha_i \times \left[(1-\epsilon_i) - y_i(W^TX_i + W_0)\right] + \sum \beta_i \times [-\epsilon_i]$$
$$\bullet \frac{1}{2}\sum\sum \alpha_i\alpha_j y_i y_j X_i X_j + C\sum \epsilon_i + \sum \alpha_i$$
$$\bullet - \sum \alpha_i \epsilon_i - \sum\sum \alpha_i\alpha_j y_i y_j X_i X_j - \sum \alpha_i y_i W_0 - \sum \beta_i \epsilon_i$$
$$\bullet - \frac{1}{2}\sum\sum \alpha_i\alpha_j y_i y_j X_i X_j + C\sum \epsilon_i + \sum \alpha_i - \sum \alpha_i\epsilon_i - W_0\sum \alpha_i y_i - \sum \beta_i \epsilon_i$$
$$\bullet W_0 \sum \alpha_i y_i = 0 \quad \text{و} \quad C = \alpha_i + \beta_i \quad \text{از فرمول های قبل بدست می‌آیند}$$
$$\bullet - \frac{1}{2}\sum\sum \alpha_i\alpha_j y_i y_j X_i X_j + \sum \alpha_i\epsilon_i + \sum \beta_i\epsilon_i + \sum \alpha_i - \sum \alpha_i\epsilon_i - \sum \beta_i \epsilon_i$$
$$\Rightarrow g(\alpha) = \sum \alpha_i - \frac{1}{2}\sum\sum \alpha_i\alpha_j y_i y_j X_i X_j$$

شبیه به همان حالت قبلی شد و تعداد پارامتر های دوگان آن با سوال ۲ که SVM معمولی بود یکی شد، مسئله دوگان آن بصورت زیر است :

$$Dual : \max g(\alpha)$$
$$s.t$$
$$0 \le \alpha_i \le C$$
$$\sum \alpha_i y_i = 0$$

فرض کنیم مسئله دوگان را حل کردیم و به بهین آن رسیدیم یعنی $\alpha^\star$ را بدست آوردیم. با استفاده از آن مانند قبل (سوال قبل)، $W^\star = \sum \alpha_i^\star y_i X_i$ را بدست می‌آوریم.

با بررسی شرایط KKT ، مقدار $W_0^\star$ را به صورت میانگینی از آن‌ها بدست می‌آوریم. مثلا می‌گویم اگر $\alpha_i \neq 0$ بود آنگاه f_i متناظر آن باید ۰ باشد و در این مسئله که W_0 برای ما مهم است، f_i را برابر $(1-\epsilon_i^\star) - y_i(W^{\star T}X_i + W_0)$ قرار می‌دهیم و با حل این معادله $f_i = 0$ مقادیر W_0 را برای همه آن‌هایی که α_i مخالف صفر دارند بدست آورده و در نهایت $W_0^\star$ را میانگین این W_0 را در نظر می‌گیریم.

اما برای بدست آوردن $W_0^\star$ علاوه بر بهین $W^\star$ به $\epsilon_i^\star$ نیز نیاز داریم.

بطور کلی با توجه به شرط KKT اگر α_i ای غیر صفر شود، f_i آن صفر می‌شود.

- اگر α_i صفر شود ، $y_i(W^TX_i + W_0) \ge 1$
- اگر α_i مقدار C شود، $y_i(W^TX_i + W_0) \le 1$
- و اگر $0 < \alpha_i < C$ آنگاه $y_i(W^TX_i + W_0) = 1$

با استفاده از این W_0 را برابر $y_i - W^TX_i$ قرار می‌دهیم.

برای label دادن به نقاط می‌توان از تابع $f(x) = sign(W^Tx + W_0)$ استفاده کرد.

رفرنس این سوال