به نام خدا

دانشگاه تهران
پردیس دانشکده‌های فنی
دانشکده برق و کامپیوتر

شناسائی الگو

تمرین شماره ۵

سجاد پاکدامن ساوجی
۸۱۰۱۹۵۵۱۷

فهرست

چکیده

در این تمرین با مباحث support vector machines بیشتر از پیش آشنا شدیم. در سوالات ابتدایی با نحوه‌ی ساخت کرنل (هسته های جدید) از روی هسته های موجود آشنا شدیم و در سوالات پایانی هم به پیاده سازی و استفاده از این طبقه بند با کرنل خطی و RBF پرداختیم.

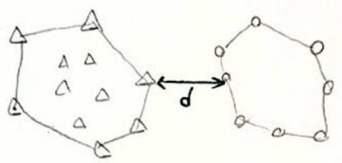
در تمرین پایانی در ابتدا مشکل هسته هایی با درجه فضای بی‌نهایت مانند RBF را متوجه شدیم که به علت درجه آزادی بیش‌از حدی که دارند به سادگی روی داده ها overfit می‌شوند. در ادامه اهمیت پارامتر های مدل svm را دیدیم و متوجه شدیم که با یک grid search ساده می‌توان مدل را fine tune کرد و به دقت‌های بسیار بالاتری رسید.

در سوال پایانی اهمیت normalization را در داده ها مشاهده کردیم که با یک normalization ساده می‌توان به صورت قابل توجهی عملکرد طبقه بند را افزایش داد.

سوال 1

problem 1) $\mathrm{conv}[X^+] = \sum_{i: y_i=1} \lambda_i x_i \qquad \mathrm{Conv}[X^-] = \sum_{j: y_j=-1} \lambda_j x_j$

$$\min_{\underline{\lambda}} \left\| \sum_i \lambda_i x_i - \sum_j \lambda_j x_j \right\|^2 \quad \text{s.t:} \ \sum_i \lambda_i = 1 \ , \ \sum_j \lambda_j = 1 \ , \ \lambda_i \geq 0$$



$$\left\| \sum_i \lambda_i x_i - \sum_j \lambda_j x_j \right\|^2 = \left(\sum_i \lambda_i x_i - \sum_j \lambda_j x_j \right)^T \left(\sum_i \lambda_i x_i - \sum_j \lambda_j x_j \right)$$

$$= \left(\sum_i \lambda_i x_i^T - \sum_j \lambda_j x_j^T \right) \left(\sum_i \lambda_i x_i - \sum_j \lambda_j x_j \right)$$

$$= \sum_i \sum_{i'} \lambda_i \lambda_{i'} \underbrace{x_i^T x_{i'}}_{A} + \sum_j \sum_{j'} \lambda_j \lambda_{j'} \underbrace{x_j^T x_{j'}}_{B} - 2 \sum_i \sum_j \lambda_i \lambda_j \underbrace{x_i^T x_j}_{P}$$

$$= \begin{bmatrix} \Lambda_i \\ \Lambda_j \end{bmatrix}^T \begin{bmatrix} \underline{A} & -\underline{P} \\ -\underline{P} & \underline{B} \end{bmatrix} \begin{bmatrix} \Lambda_i \\ \Lambda_j \end{bmatrix} = \Lambda^T \underline{D} \Lambda$$

minimize $\Lambda^T \underline{D} \Lambda$ s.t $\Lambda^T \underline{1} = 2$, $\Lambda > 0$

$$\sum_i \sum_{i'} \underbrace{y_i y_{i'}}_{1} \lambda_i \lambda_{i'} x_i^T x_{i'} + \sum_j \sum_{j'} \underbrace{y_j y_{j'}}_{1} \lambda_j \lambda_{j'} x_j^T x_{j'} + \sum_i \sum_j \underbrace{y_i y_j}_{-1} x_i^T x_j$$

$$+ \sum_j \sum_i \underbrace{y_i y_j}_{-1} x_i^T x_j = \sum_{i,j} y_i y_j x_i x_j \lambda_i \lambda_j$$

minimize $\sum_{i,j} y_i y_j x_i^T x_j \lambda_i \lambda_j$ s.t $\sum_i \lambda_i = 1$, $\sum_j \lambda_j = 1$

$$\equiv \sum_{k=1}^{N} \lambda_k = 2 \ , \ \sum_{k=1}^{N} y_k \lambda_k = 0$$

$$, \Lambda > 0$$

optimize $\sum_{i,j} y_i y_j x_i^T x_j \lambda_i \lambda_j$ s.t $\lambda > 0$ $1^T \lambda = 2$ $\Lambda^T Y = 0$

$\Rightarrow$ optimize $2 - \Lambda^T \underline{\underline{D}} \Lambda$ = optmize $1^T \Lambda - \frac{1}{2} \Lambda^T \underline{\underline{D}} \Lambda$

s.t $1 \Lambda^T = 1$, $\Lambda^T Y = 0$, $\Lambda \geq 0$

* which is duel problem of hard margin SVM

سوال 2

Problem 2)

A) $k(u,v) = \alpha k_1(u,v) - \beta k_2(u,v) \quad , \alpha, \beta \geq 0$

is invalid. consider the case $\alpha = 1 , \beta = 1 \; , k_2(u,v) = k_1(u,v) \Rightarrow k(u,v) = 0$

which is not PD.

B) $K(u,v) = k_1(u,v)\, k_2(u,v)$

$$k_1(u,v) = \langle \varphi_1(u), \varphi_1(v) \rangle = \varphi_1^T(u)\, \varphi_1(v)$$

$$k_2(u,v) = \langle \varphi_2(u), \varphi_2(v) \rangle = \varphi_2^T(u)\, \varphi_2(v)$$

$$k_1(u,v)\, k_2(u,v) = \varphi_1^T(u)\varphi_1(v) \times \varphi_2^T(u)\varphi_2(v)$$

$$\varphi_1(u) = [a_1(u) \; \cdots \; a_m(u)]^T \qquad \varphi_2(u) = [b_1(u) \; \cdots \; b_m(u)]^T$$

$$\varphi_1^T(u)\varphi_1(v) = \sum_{i=1}^{m} a_i(u) a_i(v) \quad , \quad \varphi_2^T(u)\varphi_2(v) = \sum_{j=1}^{N} b_j(u) b_j(v)$$

$$\varphi_1^T(u)\varphi_1(v) \times \varphi_2^T(u)\varphi_2(v) = \sum_{j=1}^{N}\sum_{i=1}^{M} a_i(u) a_i(v)\, b_j(u)\, b_j(v)$$

$$= \sum_{j=1}^{N}\sum_{i=1}^{M} \underbrace{a_i(u)\, b_j(u)}_{C_{ij}(u)} \times \underbrace{a_i(v)\, b_j(v)}_{C_{ij}(v)} = \sum_{j=1}^{N}\sum_{i=1}^{M} C_{ij}(u)\, C_{ij}(v) \rightarrow \text{inner product}$$

C) $K(u,v) = \exp\left\{\frac{-\|u-v\|^2}{\sigma^2}\right\} = \exp\left\{-\frac{1}{\sigma^2}\left[(u-v)^T(u-v)\right]\right\}$

$$= e^{xp}\left\{-\frac{1}{\sigma^2}\left[(u^T * u) + (v^T v) - 2 u^T v\right]\right\}$$

$$= \underbrace{\exp\left\{-\frac{1}{\sigma^2}(u^T u + v^T v)\right\}}_{C_0} \exp\left\{\underbrace{\frac{2}{\sigma^2}}_{C_1} u^T v\right\}$$

$$\xrightarrow{\text{taylor expansion}} C_0 \sum_{i=1}^{\infty} \frac{1}{i!} (C_1 u^T v)^i$$

$\rightarrow$ Product of kernels and sum of kernels are kernel ✓

D) $K(u,v) = g(u)g(v)$ mercer's theorem

$$\iint K(u,v)g(u)g(v)\,du\,dv = \iint g^2(u)g^2(v)\,du\,dv = \underbrace{\int g^2(u)\,du}_{\geq 0} \times \underbrace{\int g^2(v)\,dv}_{\geq 0}$$

if $g(u), g(v) \in L^2 \Rightarrow K(u,v)$ is mercer kernel

E) $K(u,v) = f(k_1(u,v))$

$$\underset{\text{polynomial}}{f_1(x)} = \sum_{i=1}^{\infty} a_i x^i \Rightarrow f(k_1(u,v)) = \sum_{i=1}^{\infty} a_i\, k_1^i(u,v)$$

product of kernels are kernel $\Rightarrow k_1^i(u,v) \longrightarrow$ is kernel

sum of kernels with positive factors is kernel

$$\Rightarrow \sum_{i=1}^{\infty} a_i\, k^i(u,v) \longrightarrow \text{is kernel}$$

سوال 3

problem 3)

1. relax conditions $y_i(w^T x_i + b) \geq 1 - \varepsilon_i$, $\varepsilon_i \geq 0$

minimize $\frac{1}{2} w^T w$ s.t: $y_i(w^T x_i + b) \geq 1 - \varepsilon_i$, $\varepsilon_i \geq 0$

2. $\varepsilon_i = 0 \Rightarrow y_i(w^T x_i + b) \geq 1$

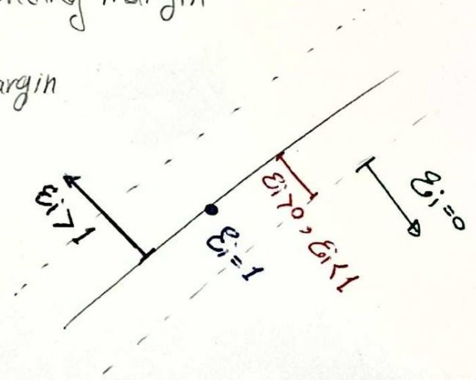
eigther on margin or correct classified

$\varepsilon_i = 1 \Rightarrow y_i(w^T x_i + b) \geq 0$

eigther on the hyper plane or in the corresponding margin

$0 < \varepsilon_i < 1 \Rightarrow$ data point is in the margin

$\varepsilon_i > 1 \Rightarrow$ miss classified



سوال 4

problem 4) $J(\underline{w},\varepsilon) = \frac{1}{2} w^T w + \frac{1}{2} \sum_{i=1}^{l} \varepsilon_i^2$

$L(\underline{w}, \underline{\varepsilon}, \underline{\Lambda}, \underline{R}, b) = \frac{1}{2} w^T w + \frac{1}{2} \sum_{i=1}^{l} \varepsilon_i^2 - \sum_{i=1}^{l} \alpha_i [y_i (x_i^T w + b) - 1]$

$- \sum_{i=1}^{l} \varepsilon_i (\alpha_i + r_i)$

$\frac{\partial L}{\partial w} = 0 \Rightarrow \underline{w} = \sum_{i=1}^{l} \alpha_i y_i x_i \qquad \frac{\partial L}{\partial b} = 0 \Rightarrow \Lambda^T Y = 0$

$\frac{\partial L}{\partial \varepsilon_i} = 0 \Rightarrow c\varepsilon_i = r_i + \alpha_i \Rightarrow c\underline{\varepsilon} = \underline{R} + \underline{\Lambda}$

$L(\underline{w}, \underline{\varepsilon}, \Lambda, R) = 1^T \Lambda - \frac{1}{2} \Lambda^T \underline{\underline{D}} \Lambda + \frac{c}{2} \sum_{i=1}^{l} \varepsilon_i^2 - c \sum_{i=1}^{l} \varepsilon_i^2$

$L(\Lambda, \underline{\varepsilon}) = 1^T \Lambda - \frac{1}{2} \Lambda^T \underline{\underline{D}} \Lambda - \frac{c}{2} \sum_{i=1}^{l} \varepsilon_i^2$

only ε_i^0 contributes to the sum $\sum_{i=1}^{l} \varepsilon_i^2$ and from KKT $\rightarrow \varepsilon_i r_i = 0$

$\Rightarrow$ if $\varepsilon_i \neq 0 \rightarrow r_i = 0 \Rightarrow \varepsilon_i = \alpha_i \Rightarrow$

$L(\underline{\Lambda}, \underline{\varepsilon}) = 1^T \Lambda - \frac{1}{2} \Lambda^T D \Lambda - \frac{c}{2} \Lambda^T \Lambda$

$L(\Lambda) = 1^T \Lambda - \frac{1}{2} \Lambda^T \underbrace{[D + cI]}_{\underline{\underline{D'}}} \Lambda \quad \text{s.t } \Lambda \geq 0, \Lambda^T Y = 0$

$L(\Lambda) = 1^T \Lambda - \frac{1}{2} \Lambda^T \underline{\underline{D'}} \Lambda \quad \text{s.t } \Lambda \geq 0, \Lambda^T Y = 0$

$\hookrightarrow$ just like hard margin optimization

thus it can be interpreted as linearly seprable case.

سوال 5

دقت شود که پیاده سازی‌های خواسته شده در این سوال ، در jupyter notebook است که به همراه گزارش ارسال شده است.

۱. در قسمت اول برای جداسازی داده های آموزش از داده‌های آزمایش از تابع train_test_split استفاده کردشد. سپس طبقه بند svm را بدون مشخص کردن پارامتر خاصی روی داده های آموزش داده شد. این طبقه بند بر روی داده های آموزش به دقت ۱۰۰ رسید در حالی که روی داده های آزمایش دقت آن ۴۲ بود. با توجه به نتایج بدست آمده این تحلیل ارائه می‌شود که به دلیل درجه آزادی بیش از حد طبقه بند روی داده های آموزش بیش‌برازش یا همان overfit کرده است.

۲. در قسمت بعدی از یک طبقه بند svm با کرنل خطی استفاده شد. این طبقه‌بند روی داده های آموزش به دقت ۳۸ درصد و روی داده‌های آزمایش به دقت ۳۱ درصد رسید. با توجه به نتایج بدست آمده این تحلیل ارائه می‌شود که طبقه‌بند بیش‌برازش انجام نداده است و علت خطای آن به دلیل ساختار با درجه آزادی کم با توجه با داده ها می‌باشد . یا به عبارتی این داده ها به صورت خطی جداپذیر نیستند.

۳. در این قسمت خواسته شده است که با استفاده از normalization طبقه بند هارا بهبود ببخشیم. در مورد طبقه بند svm با کرنل RBF ، روی داده های آموزش به دقت ۶۶ درصد و روی داده های آزمایش به دقت ۶۰ درصد رسید. مشاهده می‌شود که به علت عملیات feature conditioning دیگر مدل overfit نکرده است. برای svm با کرنل خطی نیز عملیات را انجام می‌دهیم و این طبقه‌بند روی داده های آموزش به دقت ۷۵ درصد و روی داده های آزمایش به دقت ۷۲ درصد رسید. مشاهده می‌شود که عملکرد linear svm نیز بهبود یافته است.

۴. در پایان بر روی داده های normalized شده با استفاده از gridsearch سعی کردیم که طبقه‌بند svm را با هسته rbf به پارامتر های بهینه برسانیم یا همان fine tune کنیم. مقادیر بهینه در زیر آمده اند

C = 287.5

gamma = 0.25

با این پارامتر ها طبقه‌بند به دقت ۹۴ درصد روی داده های آموزش رسید و روی داده های آزمایش به دقت ۷۷ رسید.

پیوست 1: روند اجرای برنامه

پیاده سازی های سوال ۵ در jupyter notebook به همراه گزارش ارسال شده است.