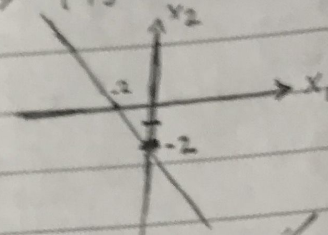
مسئله ۱. الف)

$(w_0, w_1, w_2) = (2, 1, 1) \rightarrow \begin{cases} w = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ b = 2 \end{cases}$

Decision Boundary: $\{x \mid w^T x + b = 0\} \Rightarrow \begin{bmatrix} 1 \\ 1 \end{bmatrix}^T x + 2 = 0$

$\Rightarrow x_1 + x_2 = -2 \Rightarrow x_2 = -x_1 - 2 \Rightarrow$



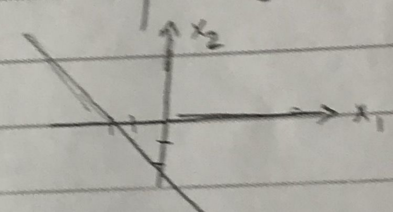
ب) الگوریتم پرسپترون هنگام دیدن یک داده بدین صورت عمل می‌کند که درصورتی که این داده اشتباه دسته‌بندی شده باشد ($y_i a_i < 0$)، w و b را به این صورت آپدیت می‌کند:

$\begin{cases} w \leftarrow w + y_i x_i \\ b \leftarrow b + y_i \end{cases}$ داده i ام

← داده (i+1) ام :

$a_{i+1} = w^T x_{i+1} + b = 2 + 1 + 2 = 5$

$\Rightarrow y_{i+1} a_{i+1} > 0 \Rightarrow$ داده به درستی طبقه‌بندی می‌شود، در نتیجه تغییری اعمال نمی‌شود و مرز تصمیم ثابت می‌ماند



← داده (i+2) ام :

$a_{i+2} = w^T x_{i+2} + b = 3 + 2 + 2 = 7$

$\Rightarrow y_{i+2} a_{i+2} = -7 < 0 \Rightarrow$ update w, b

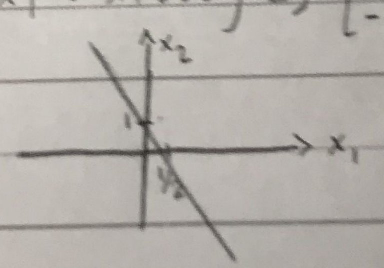
$\Rightarrow 1)\ w := w + y_{i+2} x_{i+2} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - 1 \times \begin{bmatrix} 3 \\ 2 \end{bmatrix}$

$\Rightarrow w_{new} = \begin{bmatrix} -2 \\ -1 \end{bmatrix}$

$\Rightarrow 2)\ b := b + y_{i+2} = 2 - 1 = 1 \Rightarrow b_{new} = 1$

$\Rightarrow$ مرز تصمیم : $\{x \mid w^T x + b = 0\} \Rightarrow \begin{bmatrix} -2 \\ -1 \end{bmatrix}^T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + 1 = 0 \Rightarrow x_2 = -2x_1 + 1$

$\Rightarrow$



$w = \begin{bmatrix} -2 \\ -1 \end{bmatrix}$

$b = 1$

Subject :   Date

مسئله ۲. ماتریس داده‌ها و لیبل‌ها را بدین صورت در نظر می‌گیریم:

$$X = \begin{bmatrix} 1 & 1/2 \\ -1 & -1 \end{bmatrix} \begin{array}{l} \rightarrow \text{داده اول} \\ \rightarrow \text{داده دوم} \end{array} \quad \text{و} \quad y = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad \text{و} \quad W^{(0)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{و} \quad b^{(0)} = 0$$

* یادگیری ابتدا با x_1 سپس با x_2: داده x_1 ←

$$a_1 = w^{(0)T} x_1 + b^{(0)} = 0$$

$$\Rightarrow y_1 a_1 \leq 0 \Rightarrow \text{آپدیت}: \begin{cases} w^{(1)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 1/2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1/2 \end{bmatrix} \\ b^{(1)} = 0 + 1 = 1 \end{cases}$$

داده x_2 ←

$$a_2 = w^{(1)T} x_2 + b^{(1)} = -1 - \tfrac{1}{2} + 1 = -\tfrac{1}{2}$$

$\Rightarrow y_2 a_2 > 0 \Rightarrow$ تغییر نمی‌کند $\Rightarrow$ داده x_1: $(w^{(1)T} x_1 + b^{(1)}) \times y_1 > 0 \Rightarrow$ خاتمه

$\Rightarrow$ بردارهای نهایی: $w = \begin{bmatrix} 1 \\ 1/2 \end{bmatrix}$ و $b = 1$

* یادگیری ابتدا با x_2، سپس با x_1: داده x_2 ←

$$a_2 = w^{(0)T} x_2 + b^{(0)} = 0$$

$$\Rightarrow y_2 a_2 \leq 0 \rightarrow \text{آپدیت}: \begin{cases} w^{(1)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} - \begin{bmatrix} -1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ b^{(1)} = 0 - 1 = -1 \end{cases}$$

داده x_1 ←

$$a_1 = w^{(1)T} x_1 + b^{(1)} = 1 + \tfrac{1}{2} - 1 = \tfrac{1}{2}$$

$\Rightarrow y_1 a_1 > 0 \Rightarrow$ تغییر نمی‌کند $\Rightarrow$ داده x_2: $(\underbrace{w^{(1)T} x_2 + b^{(1)}}_{-3}) \underbrace{y}_{-1} > 0$

$\Rightarrow$ خاتمه $\Leftarrow$ بردارهای نهایی: $w = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ و $b = -1$

Date ....

Subject :

مسئله ۳. بردارهای وزن را بدین صورت در نظر می‌گیریم: $w^{(0)}$ بردار اولیه و صفر است.

$w^{(1)}$ بردار وزن بعد از اولین آپدیت در الگوریتم پرسپترون، $w^{(k)}$ بردار وزن بعد از

(k)امین آپدیت. فرض می‌کنیم متغیر b داخل w هست. همچنین با توجه به آنکه

$w^{(i)}$ را برای iامین آپدیت تعریف کردیم، در نتیجه: (I) $y\, w^{(i)^T} x < 0$

حال با استقرا نشان می‌دهیم: (A) $w^{*^T} w^{(k)} \geq k\gamma$

اثبات با استقرا: فرض می‌کنیم به ازای k درست باشد. حال برای (k+1)

اثبات می‌کنیم. می‌دانیم: $w^{*^T} w^{(k)} \geq k\gamma$

فرض می‌کنیم $w^{(k+1)}$ w یک (x,y) آپدیت شده باشد: $w^{*^T} w^{(k+1)} = w^{*^T}(w^{(k)} + yx) = w^{*^T} w^{(k)} + \underbrace{y w^{*^T} x}_{\gamma}$

طبق استقرا $\Longrightarrow$ $w^{*^T} w^{(k+1)} \geq k\gamma + \gamma \Rightarrow w^{*^T} w^{(k+1)} \geq (k+1)\gamma$

حکم اثبات شد.

حال با استقرا نشان می‌دهیم: (B) $\|w^{(k)}\|^2 \leq kR^2$

اثبات با استقرا: فرض می‌کنیم برای n=k درست است: $\|w^{(k)}\|^2 \leq k$

حال برای n=k+1 داریم: $\|w^{(k+1)}\|^2 = \|w^{(k)} + yx\|^2$

$= \|w^{(k)}\|^2 + y^2\|x\|^2 + 2y\, w^{(k)^T} x$

MICRO

Date

Subject :

حال با توجه به آنکه تمام داده‌ها در ابرکره‌ای با شعاع R قرار دارند داریم: $\|x\|^2 \leq R^2$

و چون لیبل‌ها ۱-، ۱+ هستند ، $y^2=1$ و طبق (I) $y w^{(k)^T} x \leq 0$

$$\Longrightarrow \|w^{(k+1)}\|^2 = \|w^{(k)}\|^2 + y^2\|x\|^2 + \underbrace{2y\, w^{(k)^T} x}_{\leq 0}$$

$$\Rightarrow \|w^{(k+1)}\|^2 \leq \underbrace{\|w^{(k)}\|^2}_{\text{استقرا}} + \underbrace{y^2\|x\|^2}_{\leq R^2}$$

$$\Rightarrow \|w^{(k+1)}\|^2 \leq kR^2 + R^2 \Rightarrow \|w^{(k+1)}\|^2 \leq (k+1)R^2$$

حکم ثابت شد

$$\Rightarrow A: w^{*^T} w^{(k)} \geq k\gamma \rightarrow \|w^*\|\|w^{(k)}\| \cos\theta \geq k\gamma$$

$$\Rightarrow \|w^{(k)}\| \geq \frac{k\gamma}{\|w^*\|\cos\theta}$$

$$\overset{B}{\Rightarrow} \frac{k\gamma}{\|w^*\|\cos\theta} \leq \|w^{(k)}\| \leq \sqrt{k}R$$

$$\Rightarrow \cos\theta \geq \frac{\sqrt{k}\,\gamma}{\|w^*\| R} \Longrightarrow$$

افزایش k ، منجر می‌شود که سمت راست به یک میل کند

و چون $\cos\theta \leq 1$ ، داریم:

$$\frac{\sqrt{k}\,\gamma}{\|w^*\| R} = 1 \Rightarrow k = \frac{\|w^*\|_2^2 R^2}{\gamma^2}$$

MICRO

مسئله ۴. آ) ادعا می‌کنیم که در جواب بهینه $\forall i : \xi_i \geq 0$

اثبات: برهان خلف: فرض خلف: در جواب بهینه حداقل یک ξ_i وجود دارد

که $\xi_i < 0$. حال داریم: $\xi_i < 0 \rightarrow 1 - \xi_i > 1$

اگر $\xi_i = 0$ بگذاریم هم قید مسئله باز $\Leftarrow \; y_i(w^T x_i + b) \geq 1 \; \Rightarrow$

برقرار خواهد بود و مشکلی بوجود نمی‌آید

حال با توجه به آنکه مجذور ξ در objective function ظاهر شده است

در نتیجه مقدار مثبتی را به آن اضافه می‌کند در حالی که اگر $\xi_i = 0$ (نشان دادیم

که قیود مسئله را برقرار می‌کند)، مقدار objective function کمتر می‌شد

که این در تضاد بهینه بودن این جواب است. در نتیجه حکم اولیه درست

است و در این صورت بندی تضمین می‌شود که $\forall i : \xi_i \geq 0$

ب) $$\mathcal{L}(w, b, \xi, \alpha) = \frac{1}{2} \underbrace{\|w\|^2}_{w^T w} + \frac{c}{2} \sum_{i=1}^{N} \xi_i^2 - \sum_{i=1}^{N} \alpha_i \left(y_i (w^T x_i + b) - 1 + \xi_i \right)$$

$$\frac{\partial \mathcal{L}}{\partial w} = \frac{1}{2} \frac{\partial}{\partial w} w^T w - \sum \frac{\partial}{\partial w} \left(\alpha_i y_i (w^T x_i) \right) = 0$$

$$= w - \sum_{i=1}^{N} \alpha_i y_i x_i = 0 \Rightarrow w = \sum_{i=1}^{N} \alpha_i y_i x_i$$

$$\Rightarrow \frac{\partial \mathcal{L}}{\partial b} = - \sum_{i=1}^{N} \frac{\partial}{\partial b} \alpha_i y_i b = - \sum \alpha_i y_i = 0 \Rightarrow \sum_{i=1}^{N} \alpha_i y_i = 0$$

Subject : Date ....

$$= \frac{\partial \mathcal{L}}{\partial \xi} = \begin{bmatrix} \frac{\partial \mathcal{L}}{\partial \xi_1} \\ \vdots \\ \frac{\partial \mathcal{L}}{\partial \xi_i} \\ \vdots \\ \frac{\partial \mathcal{L}}{\partial \xi_N} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \end{bmatrix} \Rightarrow \frac{\partial \mathcal{L}}{\partial \xi_i} = c\xi_i - \alpha_i = 0 \Rightarrow c\xi_i = \alpha_i$$

$$c \underbrace{\begin{bmatrix} \xi_1 \\ \vdots \\ \xi_n \end{bmatrix}}_{\xi} = \underbrace{\begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}}_{\alpha}$$

ج) فرم عبارت dual عبارت است از : $g^* = \max\limits_{\alpha} \min\limits_{w, b, \xi} \mathcal{L}(w, b, \xi, \alpha)$

قسمت کمینه کردن را در قسمت ب انجام دادیم. در نتیجه نتایج آن را

در لاگرانژین جایگذاری می‌کنیم: $\min\limits_{w,b,\xi} \mathcal{L}(w, b, \xi, \alpha)$

$$= \frac{1}{2} \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j x_i^T x_j + \frac{1}{2} \sum_{i=1}^{N} \frac{\alpha_i}{\xi_i} \xi_i^2$$

$$- \sum_{i=1}^{N} \alpha_i \left(y_i \left(\left(\sum_{j=1}^{N} \alpha_j y_j x_j^T \right) x_i + b \right) - 1 + \xi_i \right)$$

$$= -\frac{1}{2} \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j x_i^T x_j - \frac{1}{2} \sum_{i=1}^{N} \alpha_i \xi_i - \underbrace{\cancel{\sum_{i=1}^{N} \alpha_i y_i} b}_{0} + \sum_{i=1}^{N} \alpha_i$$

$$= -\frac{1}{2} \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j x_i^T x_j - \frac{1}{2} \sum_{i=1}^{N} \frac{\alpha_i^2}{c} + \sum_{i=1}^{N} \alpha_i$$

$$\Rightarrow \text{Dual: } \max_{\alpha} \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j x_i^T x_j - \frac{1}{2} \sum_{i=1}^{N} \frac{\alpha_i^2}{c}$$

$$\text{s.t.} \quad \alpha_i \geq 0 \quad \forall i = 1, \dots, N$$

$$\sum_{i=1}^{N} \alpha_i y_i = 0$$

Subject : Date

سوال ۵. کرنل RBF عبارت است از:

$$K(x_1,x_2)=e^{-\frac{\|x_1-x_2\|^2}{2\sigma^2}}$$

$\sigma=1$ در نظر می‌گیریم که در حل مسئله مشکلی پیش نمی‌آید

$$K(x_1,x_2)=e^{-\frac{\|x_1-x_2\|^2}{2}}$$

حال با توجه به اینکه x_1-x_2 در عبارت کرنل ظاهر شده است تعریف

می‌کنیم: $K(x_1,x_2)=f(x-x_2)$ و حال $x_1-x_2=t$ در نتیجه داریم:

$$K(x_1,x_2)=f(x_1-x_2)=f(t)=e^{-t^2/2}$$

حال می‌خواهیم از خواص MGF و تابع مشخصه توزیع نرمال استاندارد

استفاده کنیم. در حالت حقیقی می‌دانستیم که

$$E[e^{tZ}]=e^{\frac{1}{2}t^2}$$

$$Z\sim N(0,1)$$

در حالت موهومی داریم:

$$E[e^{itZ}]=e^{-t^2/2}$$

که منفی به علت $i^2=-1$ بوجود می‌آید. حال از این نتیجه استفاده می‌کنیم:

$$\sum_{j=1}^{N}\sum_{k=1}^{N}a_ja_k f(x_j-x_k)=\sum_{j=1}^{N}\sum_{k=1}^{N}a_ja_k E[e^{i(x_j-x_k)Z}]$$

$$\overset{\text{خطی بودن امید ریاضی}}{=} E\left[\sum\sum a_j e^{ix_jZ}a_k e^{-ix_kZ}\right]$$

$$=E\Big[\underbrace{\Big(\sum a_j e^{ix_jZ}\Big)^2}_{\geq 0}\Big]\Rightarrow \geq 0 \Rightarrow$$

در نتیجه مثبت نیمه معین است

مسئله ششم. ۱. این کرنل همان ضرب داخلی عادی است که تنها می تواند داده‌های خطی جداپذیر را از هم تفکیک کند، چرا که از مرتبه اول است، در نتیجه تنها دیتاست (آ) را می تواند جدا کند.

۲. به طور کلی feature map مربوط به کرنل RBF، یک فضای بی‌نهایت بعدی می‌سازد که توانایی جداسازی هر دو دسته‌ای را در فضای بی‌نهایت بعدی دارد. اما در فضای فعلی دیتاست به طور کلی کرنل RBF به ازای هر داده یک توزیع گاوسی فیت می‌کند و هر چه γ در $e^{-\gamma\|x_1-x_2\|^2}$ بزرگتر باشد، توزیع گاوسی فیت شده خاص‌تر می‌شود، به عبارتی تنها داده‌هایی که واقعا نزدیک به هم باشند کرنل بالایی می‌گیرند

← با این تفاسیر این کرنل تمام دیتاست‌های (آ) (ب) (ج) (د) را تفکیک می‌کند ((تنها در دیتاست (ب) ممکن است مشکل ایجاد شود چرا که تلاقی زیادی وجود دارد))

۳. این کرنل همان کرنل polynomial یا چندجمله‌ای است که توانایی تفکیک داده‌هایی که به صورت چند جمله‌ای جدا شده‌اند را دارد. در نتیجه (آ) و (د) را به راحتی تفکیک می‌کند. (ب) را نمی تواند تفکیک کند. (ج) را هم به نظر به علت یک datapoint مرزی مانع از تفکیک به صورت دایره شود

۴. توضیحات این بخش همانند قسمت ۲ است، اما چون γ کم است به نقاطی که کمتر شبیه هستند را مقدار بالایی نسبت می‌دهد. در نتیجه (ب) را نمی‌تواند تفکیک کند اما (آ) و (ج) و (د) را می‌تواند تفکیک کند

Subject : مسئله ی ۷ — Date

۱. ماتریس گرم مربوط به k_3 را تشکیل می دهیم:

$$G_3 = [k_3(x_i, x_j)]_{ij} = [k_1(x_i, x_j) + k_2(x_i, x_j)]_{ij}$$

$$= [k_1(x_i, x_j)]_{ij} + [k_2(x_i, x_j)]_{ij} = G_1 + G_2$$

حال می خواهیم از تعریف مثبت نیمه معین بودن G_3 برای اثبات معتبر بودن استفاده کنیم

$$\Rightarrow u^T G_3 u = u^T(G_1 + G_2)u = u^T G_1 u + u^T G_2 u$$

k_1 و k_2 معتبر $\Rightarrow \begin{cases} u^T G_1 u \geq 0 \\ u^T G_2 u \geq 0 \end{cases} \Rightarrow u^T G_3 u \geq 0 \Leftrightarrow G_3$ مثبت نیمه معین

$\leftarrow k_3$ معتبر است. همچنین این حکم به ازای تمام ترکیب خطی های کرنل های معتبر با ضرایب مثبت به همین ترتیب اثبات می شود و تعمیم می پذیرد.

۲. می دانیم یک کرنل معتبر است اگر به ازای هر x, y داشته باشیم:

$$k(x,y) = \phi(x)^T \phi(y)$$

حال برای k_4 داریم:

$$k_4(x,y) = k_1(x,y)\, k_2(x,y)$$

$$\Rightarrow k_1, k_2 \text{ معتبر} \Rightarrow k_4(x,y) = \phi^{(1)}(x)^T \phi^{(1)}(y)\; \phi^{(2)}(x)^T \phi^{(2)}(y)$$

$\phi^{(1)}, \phi^{(2)}$ تبدیل های فضای مربوط به آن کرنل خاص هستند

Notation در توان تنها برای شماره گذاری است و به معنای توان نیست

$$\Rightarrow k_4(x,y) = \left(\sum_{i=1}^{N} \phi_i^{(1)}(x)\phi_i^{(1)}(y)\right) \times \left(\sum_{j=1}^{M} \phi_j^{(2)}(x)\phi_j^{(2)}(y)\right)$$

MICRO

i و j نشان دهنده مولفه های بردارهای مربوط هستند. همچنین N, M ابعاد فضای جدید هستند

Subject : Date

$$\Rightarrow k_4(x,y) = \sum_{i}^{N} \sum_{j=1}^{M} \phi_i^{(1)}(x)\, \phi_i^{(1)}(y)\, \phi_j^{(2)}(x)\, \phi_j^{(2)}(y)$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{M} \left(\phi_i^{(1)}(x)\, \phi_j^{(2)}(x)\right) \left(\phi_i^{(1)}(y)\, \phi_j^{(2)}(y)\right)$$

حال تعریف می‌کنیم $\phi_{ij}^{(3)}(\theta) = \phi_i^{(1)}(\theta)\, \phi_j^{(2)}(\theta)$ که $\phi^{(3)}(\theta)$

یک بردار $N \times M$ بعدی است (یک ماتریس NM بعدی که آن را به صورت یک

بردار بازنویسی کرده‌ایم). در نتیجه داریم: $= \sum_{i=1}^{N} \sum_{j=1}^{M} \underbrace{\phi_{ij}^{(3)}(x)\, \phi_{ij}^{(3)}(y)}_{\text{تعریف ضرب داخلی دو بردار } N\times M \text{ بعدی}}$

$= \phi^{(3)T} \phi^{(3)}$ در نتیجه توانستیم $k_4(x,y)$ را به صورت

ضرب داخلی feature map نشان دهیم

و چون این اثبات به ازای هر x, y برقرار است، در نتیجه k_4 یک کرنل معتبر است

۳. با توجه به اینکه $k_1(x_1, x_2) \in \mathbb{R}$، بسط تیلور تابع نمایی را حول نقطه

صفر می‌نویسیم: $e^0 + \frac{e^0}{1!} k_1(x_1, x_2) + \frac{e^0}{2!} k_1(x_1, x_2)^2 + \dots$

$$= 1 + k_1(x_1, x_2) + \frac{1}{2} k_1(x_1, x_2)^2 + \frac{1}{6} k_1(x_1, x_2)^3 + \dots$$

$\leftarrow k_1(x_1, x_2)^\alpha$، حاصل ضرب α بار $k_1(x_1, x_2)$ است و طبق بخش ۲

می‌دانیم که یک کرنل معتبر است. حال با ایجاد حاصل جمع کرنل‌های معتبر بوجود

می‌آید که طبق بند الف می‌دانیم حاصل آن جمع یک کرنل معتبر است

و چون روابط بالا به ازای هر x_1, x_2 برقرار است در نتیجه k_5 معتبر است

مسئله‌ی ۷

۴. با توجه به آنکه $x_1^T x_2 \in \mathbb{R}$ ، و با توجه به آنکه $k = x_1^T x_2$ یک کرنل معتبر است، ابتدا بسط تیلور $\frac{1}{1-k}$ را می‌نویسیم:

$$\frac{1}{1-k} = 1 + k + k^2 + \dots$$

حال طبق بخش دوم مسئله ۷، k^i نیز یک کرنل معتبر است و با توجه به بخش اول مسئله ۷، مجموع کرنل‌های معتبر، یک کرنل معتبر می‌سازد

مسئله‌ی ۸. الف) در الگوریتم ~~پر~~ پرسپترون، هرگاه داده i ام را به اشتباه دسته‌بندی کنیم، مقدار $y_i x_i$ را به بردار وزن اضافه می‌کنیم. حال داده i ام، α_i بار اشتباه دسته‌بندی می‌شود، در نتیجه به ازای داده i، α_i بار آپدیت انجام می‌دهیم. یعنی $\alpha_i y_i x_i$. پس در نهایت الگوریتم پرسپترون به وزن

$$W = \sum_{i=1}^{N} \alpha_i y_i x_i$$

می‌رسد (وزن ابتدایی صفر است)

رابطه بروزرسانی وزن‌ها را بدین صورت می‌توان بازنویسی کرد:

```
for i in 1:N
    for j in 1: αi
        w ← w + yi xi
```

ب) ابتدا w بدست آمده را در $\text{sign}(w^T x)$ جایگذاری می‌کنیم:

$$\text{sign}\left(\left(\sum_{i=1}^{N} \alpha_i y_i x_i \right)^T x \right) = \text{sign}\left(\sum_{i=1}^{N} \alpha_i y_i \underbrace{x_i^T x} \right)$$

حال همانطور که در رابطه بدست آمده دیده می‌شود یک ضرب داخلی دیده می‌شود که می‌توان آنرا با Kernel (حاصل تابع) جایگزین کرد و در نتیجه دیگر لازم نیست داده‌ها را به فضای جدید برده و سپس ضرب داخلی را انجام داد. در نتیجه حجم محاسبات افزایش چشمگیری نمی‌یابد. (در واقع از این نکته استفاده کردیم که $K(x,y) = \phi^T(x)\phi(y)$)

مسئله‌ی ۹ . ۱. ابتدا فضای ورودی مسئله را تعریف می‌کنیم. فضای ورودی مسئله را تمام زیرمجموعه‌های یک مجموعه‌ی ثابت D با اعضای متناهی تعریف می‌کنیم. در نتیجه واضح است که A و B زیرمجموعه‌های D هستند چرا که عضو فضای ورودی هستند. حال $K(A,B)=2^{|A\cap B|}$ شمارش همه‌ی تعداد زیرمجموعه‌های مشترک A و B است، یعنی آن‌هایی که هم زیرمجموعه‌ی A هستند هم زیرمجموعه‌ی B.

حال با توجه به شهودی که به مفهوم کرنل داریم و با توجه به فضای ورودی که تعریف کردیم، نیاز داریم تبدیل Φ را به گونه‌ای تعریف کنیم تا شرط کرنل بودن را برآورده کند یعنی $K(A,B)=\Phi^T(A)\Phi(B)$. تبدیل Φ را بدین گونه تعریف می‌کنیم که هر عضو فضا را به یک بردار $2^{|D|}$ بعدی تبدیل می‌کند. هر یک از مولفه‌های بردار حاصل بیانگر این است که آیا زیرمجموعه‌ی متناظر آن مولفه از میان زیرمجموعه‌های D، زیرمجموعه‌ی آن عضو خاص ورودی هست یا نه. به عبارت دیگر:

$$\Phi(A)_{U_i}=\begin{cases}1 & U_i\subseteq A\\ 0 & \text{o.w}\end{cases}$$

$$i=1,2,\dots,2^{|D|}$$

حال $\Phi^T(A)\Phi(B)$ نشانگر زیرمجموعه‌های مشترک A و B است چرا که در صورتی که یک زیرمجموعه مشترک باشد، مولفه متناظر آن در هر دو بردار یک است.

Subject :

Date

در نتیجه می‌توان گفت

$$K(A,B) = 2^{|A\cap B|} = \phi^T(A)\,\phi(B)$$

۲.

$$\hat{K}(A,B) = \sum_{x_1\in A}\sum_{x_2\in B} k(x_1,x_2) \rightarrow$$

می‌دانیم k معتبر است $\rightarrow k(x_1,x_2) = \langle \phi(x_1),\phi(x_2)\rangle$

$$\Rightarrow \hat{K}(A,B) = \sum_{x_1\in A}\sum_{x_2\in B} \langle \phi(x_1),\phi(x_2)\rangle$$

$$\xRightarrow[\text{ضرب داخلی}]{\text{خاصیت}} = \left\langle \sum_{x_1\in A}\phi(x_1),\ \sum_{x_2\in B}\phi(x_2)\right\rangle$$

حال تعریف می‌کنیم:

$$\hat{\phi}(A) = \sum_{x_1\in A}\phi(x_1)$$

$$\Rightarrow \hat{K}(A,B) = \langle \hat{\phi}(A),\hat{\phi}(B)\rangle = \hat{\phi}^T(A)\,\hat{\phi}(B)$$

در نتیجه $\hat{K}$ یک کرنل معتبر است