تمرین سری دوم ماشین لرنینگ

امیرمحمد عزتی 402212269

سوال ①

$$\begin{cases} a \in \mathbb{R}^+ \\ f: \mathbb{R}^d \to \mathbb{R}^p \\ K_1, K_2 \quad \text{valid Kernels}: \mathbb{R}^d \times \mathbb{R}^d \\ K_3 \quad \text{valid Kernel}: \mathbb{R}^p \times \mathbb{R}^p \end{cases}$$

آ) $K(x,z) = K_1(x,z) + K_2(x,z)$ valid ✓

$$K_1(x,z) = \phi_1(x)^T \phi_1(z)$$

$$K_2(x,z) = \phi_2(x)^T \phi_2(z)$$

$$\Rightarrow \quad K(x,z) = \phi_1(x)^T \phi_1(z) + \phi_2(x)^T \phi_2(z)$$

$$K(x,z) = \begin{bmatrix} \phi_1(x)^T \\ \phi_2(x)^T \end{bmatrix}^T \begin{bmatrix} \phi_1(z) \\ \phi_2(z) \end{bmatrix}$$

$$\xrightarrow{\text{So}} \quad \phi(x) = \begin{bmatrix} \phi_1(x) \\ \phi_2(x) \end{bmatrix}$$

پس اگر feature map بصورت بالا حاصل concate دو feature map باشد

K کرنل valid است.

ب) valid ✓

$K(x,z) = a K_1(x,z) \quad a>0$

$K(x,z) = a\, \phi_1(x)^T \phi_1(z) = \sqrt{a}\, \phi_1(x)^T \phi_1(z) \sqrt{a}$

$\xrightarrow{So} \phi(x) = \sqrt{a}\, \phi_1(x)$

پس کرنل K نیز valid است و feature map آن مانند بالا است.

ج) valid ✓

$K(x,z) = K_3(f(x), f(z))$

$= \phi_3(f(x))^T \cdot \phi_3(f(z))$

$\phi(a) = \phi_3 \circ f(x) \rightarrow$ function composition

چون $f(x)$ خروجی p بعدی دارد و K_3 یک کرنل $p\times p$ است پس K با featur map بالا یک کرنل معتبر (valid) است.

سوال ②

آ) با تعدیل w_2 ← ثابت می‌ماند

با قرار دادن w_2 بعنوان ترم رگولاریزیشن، مرز تفکیک دهی (خط جداکننده) کمتر از x_2 تاثیر می‌پذیرد. چون w_2 کاهش یافته و نزدیک صفر می‌شود. پس خط جدا کننده عمودی خواهد بود. چون با توجه به شکل داده‌های آموزش توسط خط عمودی تفکیک پذیرند. پس خطای آموزش تغییر نخواهد کرد و صفر خواهد بود.

ب) با تعدیل w_1 ← خطا افزایش می‌یابد.

با قرار دادن w_1 بعنوان ترم رگولاریزیشن، مرز تفکیک خطی افقی خواهد شد. چون w_1 نزدیک صفر شدن و تاثیر x_1 از بین می‌رود. با توجه به شکل داده‌های آموزش، خط افقی جدا کننده‌ای وجود ندارد. پس خطای آموزش افزایش می‌یابد.

ج) با تعدیل w_0 ← خطا افزایش می‌یابد.

با قرار دادن w_0 بعنوان ترم رگولاریزیشن، عرض از مبدا خط جداکننده (همان بایاس) صفر خواهد شد و خط تفکیک از مبدا مختصات لزوما عبور می‌کند. با توجه به شکل داده‌های آموزش خطی وجود ندارد که از مرکز مختصات بگذرد و بدون خطا تفکیک انجام دهد. پس خطای آموزش افزایش می‌یابد.

د) به ازای مقادیر بزرگ C، w_1 و w_2 به سمت صفر شدن حرکت می‌کنند. از طرفی چون تعداد داده‌های دو کلاس برابر است انتظار داریم احتمال یکسانی برای هر کلاس داشته باشیم. پس:

$$P(y=1|x,w) = g(w_0+w_1x_1+w_2x_2) = g(w_0) = \frac{1}{1+e^{w_0}}$$

چون مدل رگرسیون لاجیستیک داریم پس احتمال کلاس صفر برابر است با :

$$P(y=0|x,w) = 1 - P(y=1|x,w) = 1 - \frac{1}{1+e^{-w_0}}$$

پس :

$$P(y=0|x,w) = P(y=1|x,w)$$

$$1 - \frac{1}{1+e^{-w_0}} = \frac{1}{1+e^{-w_0}} \Rightarrow \boxed{w_0 = 0}$$

$$P(y=0|x,w) = P(y=1|x,w) = 0.5 = \frac{1}{2}$$

در این شرایط جمع log احتمالات نیز ماکزیمم خواهد بود و برابر است با :

$$\sum_{i=1}^{n} \log P(y_i | x_i : w_0, w_1, w_2) = n \log\left(\frac{1}{2}\right)$$

5) اگر تعدادی داده از کلاس "+" اضافه کنیم، مجموعه داده‌های آموزش‌مان unbalanced می‌شود. پس مدل می‌خواهد احتمال بیشتر به کلاس "+" دهد. چون با مقادیر بزرگ C، ترم های w_1 و w_2 صفر شده و تاثیر نخواهند داشت و مدل‌مان صرفا به w_0 وابسته خواهد بود و آن را طوری انتخاب می‌کند که احتمال کلاس با داده بیشتر، بیشتر باشد. پس مصورت زیر انتظار داریم w_0 بزرگتر از صفر باشد.

$$P(y=1|x,w) > P(y=0|x,w)$$

$$\frac{1}{1+e^{-w_0}} > 1 - \frac{1}{1+e^{-w_0}} \Rightarrow \frac{2}{1+e^{-w_0}} > 1$$

$$\Rightarrow 1+e^{-w_0} < 2 \Rightarrow e^{-w_0} < 1 \xrightarrow{\log_e} \boxed{w_0 > 0}$$

سوال ③

الف)

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

$$= e^{-\frac{x^2}{2\sigma^2}} \cdot e^{-\frac{x'^2}{2\sigma^2}} \cdot e^{\frac{xx'}{\sigma^2}}$$

حال جمله $e^{\frac{xx'}{\sigma^2}}$ را با بسط تیلور تجزیه می‌کنیم :

$$e^{\frac{xx'}{\sigma^2}} = \sum_{j=0}^{\infty} \frac{(xx')^j}{\sigma^{2j}\, j!} = 1 + \frac{xx'}{\sigma^2} + \frac{(xx')^2}{\sigma^4 \cdot 4} + \cdots$$

$$K(x,x') = e^{-\frac{x^2}{2\sigma^2}} \cdot \sum_{j=0}^{\infty} \frac{(xx')^j}{\sigma^{2j}\, j!} \cdot e^{-\frac{x'^2}{2\sigma^2}}$$

$$\begin{cases} \phi(x) = e^{-\frac{x^2}{2\sigma^2}} \left[\frac{x^0}{\sigma^0\sqrt{0!}} \quad \frac{x^1}{\sigma^1\sqrt{1!}} \quad \frac{x^2}{\sigma^2\sqrt{2!}} \quad \cdots \right] \\ \phi(x') = e^{-\frac{x'^2}{2\sigma^2}} \left[\frac{x'^0}{\sigma^0\sqrt{0!}} \quad \frac{x'^1}{\sigma^1\sqrt{1!}} \quad \frac{x'^2}{\sigma^2\sqrt{2!}} \quad \cdots \right] \end{cases}$$

چون بصورت بسط تیلور نوشته شده است پس featur map بصورت بردار ویژگی با بعد نامتناهی می‌گردد و کرنل ضرب داخلی آن ها است .

سوال 3

ب) $K(x,y) = x^T A y$

$A \in \mathbb{R}^{P \times P}$

A is symmetric

A is PSD

اگر A متقارن و PSD باشد آنگاه می‌توانیم آن تجزیه eigen آن را بصورت زیر بنویسیم .

$$A = Q \Lambda Q^T = (Q \Lambda^{1/2}) \underbrace{(Q \Lambda^{1/2})^T}_{B}$$

$$\Rightarrow A = B^T B$$

$$\Rightarrow K(x,y) = x^T B^T B y = (Bx)^T (By)$$

$$\Rightarrow \phi(x) = Bx$$

پس featur map بصورت روبه‌رو خواهد بود و K کرنل valid است .

Scanned with CamScanner