400110154 :شماره دانشجویی
سینا نمازی

۱-

آ)

$$K(x,z) = K_1(x,z) + K_2(x,z)$$

$$\Rightarrow K(x,z) = \langle \phi_1(x), \phi_1(z) \rangle + \langle \phi_2(x), \phi_2(z) \rangle = \left\langle \begin{bmatrix} \phi_1(x) \\ \phi_2(x) \end{bmatrix}, \begin{bmatrix} \phi_1(z) \\ \phi_2(z) \end{bmatrix} \right\rangle$$

$$\Rightarrow \phi(x) = \begin{bmatrix} \phi_1(x) \\ \phi_2(x) \end{bmatrix} \Rightarrow K(x,z) = \langle \phi(x), \phi(z) \rangle \Rightarrow \text{Valid kernel}$$

ب)

$$K(x,z) = a\, k_1(x,z) = a \times \langle \phi_1(x), \phi_1(z) \rangle \overset{a \geq 0}{=} \langle \sqrt{a}\,\phi_1(x), \sqrt{a}\,\phi_1(z) \rangle$$

$$\Rightarrow \phi(x) = \sqrt{a}\, \phi_1(x) \Rightarrow K(x,z) = \langle \phi(x), \phi(z) \rangle \Rightarrow \text{Valid kernel}$$

ج)

$$K(x,z) = k_3(f(x), f(z)) = \langle \phi_3(f(x)), \phi_3(f(z)) \rangle$$

$$\Rightarrow \phi(x) = \phi_3(f(x)) \Rightarrow K(x,z) = \langle \phi(x), \phi(z) \rangle \Rightarrow \text{Valid kernel}$$

آ) همچنان خطای آموزشی صفر می‌ماند، چون اگر w_2 را تعدیل کنیم و حتی آن را برابر با صفر قرار دهیم، در این حالت خط بدست آمده عمودی خواهد بود و همچنان می‌توان دو گروه داده های آموزشی را با یک خط عمودی یا یک vertical linear separator از هم جدا کرد.

ب) خطای آموزشی افزایش خواهد یافت، چون اگر w_1 را با مقادیر بالای C تعدیل کنیم، مقدار w_1^2 کمتر شده و linear separator ما افقی تر شده و با صفر شدن w_1^2، به یک خط افقی تبدیل می‌شود و نمی‌توان دو گروه داده های آموزشی را با یک horizontal linear separator از هم جدا کرد.

ج) خطای آموزشی افزایش خواهد یافت، چون اگر w_0 را با مقادیر بالای C تعدیل کنیم، مقدار w_0^2 کمتر شده و linear separator ما به مبدأ نزدیک تر شده و با صفر شدن w_0^2، به یک خط تبدیل می‌شود که از مبدأ می‌گذرد، طبق شکل داده شده، نمی‌توان دو گروه داده آموزشی را با خطی که از مبدأ عبور می‌کند از هم جدا کرد.

د) برای مقادیر بزرگ C، w_1 و w_2 به صفر میل می‌کنند:

$$w_0 = \arg\max_{w_0} \sum_{i=1}^{n} \log P(y_i \mid x_i ; w_0)$$

$$= \arg\max_{w_0} \left(8 \times \log \frac{1}{1+e^{w_0}} + 8 \times \log \frac{e^{-w_0}}{1+e^{-w_0}}\right) = \arg\max_{w_0} \left(8 \log \frac{e^{-w_0}}{(1+e^{-w_0})^2}\right)$$

$$\Rightarrow \frac{d}{dw_0} \frac{e^{-w_0}}{(1+e^{-w_0})^2} = 0 \Rightarrow w_0 = 0 \Rightarrow P(y=1 \mid x_i ; w_0) = P(y=0 \mid x_i ; w_0) = \frac{1}{2}$$

۵) مانند مورد قبل برای مقادیر بزرگ C ، w_1 و w_2 به صفر میل می‌کنند :

$$w_0 = \underset{w_0}{\arg\max} \sum_{i=1}^{n} \log P(y_i | x_i ; w_0)$$

$$= \underset{w_0}{\arg\max} \left((8+n) \times \log \frac{1}{1+e^{-w_0}} + 8 \times \log \frac{e^{-w_0}}{1+e^{-w_0}} \right)$$

$$= \underset{w_0}{\arg\max} \quad n \log \frac{1}{1+e^{-w_0}} + 8 \log \frac{e^{-w_0}}{(1+e^{-w_0})^2}$$

$\Rightarrow$ با اضافه شدن term جدید نسبت به مورد قبل $\Rightarrow$ $w_0 > 0$ $\Rightarrow$ $P(y=1 | x_i, w_0) > \frac{1}{2}$

الف)

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

$$= \exp(\langle x,x'\rangle) \cdot \exp\left(\frac{\|x\|^2}{2}\right) \cdot \exp\left(\frac{\|x'\|^2}{2}\right) =$$

$$\sum_{j=0}^{\infty} \frac{\langle x,x'\rangle^j}{\sigma^{2j} \cdot j!} \cdot \exp\left(\frac{\|x\|^2}{2}\right) \cdot \exp\left(\frac{\|x'\|^2}{2}\right) =$$

$$\sum_{j=0}^{\infty} \frac{\exp\left(\frac{\|x\|^2}{2}\right)}{\sigma^j \cdot \sqrt{j!}} \cdot \frac{\exp\left(\frac{\|x'\|^2}{2}\right)}{\sigma^j \cdot \sqrt{j!}} \cdot \langle x,x'\rangle^j = \sum_{j=0}^{\infty} C_{\sigma,j}(x) \cdot C_{\sigma,j}(x') \cdot \langle x,x'\rangle^j$$

$$\langle x,x'\rangle^j = \left(\sum_{d=1}^{n} x_d x'_d\right)^j = \sum_{k_1+k_2+\cdots+k_n=j} \binom{j}{k_1,k_2,\ldots,k_n} \prod_{d=1}^{n} (x_d x'_d)^{k_d}$$

$$= \sum_{k_1+k_2+\cdots+k_n=j} \binom{j}{k_1,k_2,\ldots,k_n}^{\frac{1}{2}} \prod_{d=1}^{n} x_d^{k_d} \cdot \binom{j}{k_1,k_2,\ldots,k_n}^{\frac{1}{2}} \prod_{d=1}^{n} {x'_d}^{k_d} = \sum_{k_1+k_2+\cdots+k_n=j} f_{j,k}(x) \cdot f_{j,k}(x')$$

$$\Rightarrow k(x,x') = \sum_{j=0}^{\infty} C_{\sigma,j}(x) \cdot C_{\sigma,j}(x') \cdot \sum_{k_1+k_2+\cdots+k_n=j} f_{j,k}(x) \cdot f_{j,k}(x')$$

$$= \sum_{j=0}^{\infty} \sum_{k_1+k_2+\cdots+k_n=j} \left(C_{\sigma,j}(x) \cdot f_{j,k}(x)\right) \cdot \left(C_{\sigma,j}(x') \cdot f_{j,k}(x')\right) = \langle \phi(x), \phi(x')\rangle$$

در اینجا هر ϕ یک وکتور با یک درایه به ازای هر ترکیبی از n (بعد هر کدام از بردارهای x و x') تا اندیس که با k نشان داده شده است که kها جمعشان برابر با j می‌شود و j هم از 0 تا ∞ است.

ب) چون A یک ماتریس نیمه مثبت معین است، می‌توان آن را به شکل روبه رو تجزیه کرد:

$$A = Q D Q^T = Q\sqrt{D}\sqrt{D}Q^T = \underbrace{Q\sqrt{D}}_{M}\sqrt{D}^T Q^T = M M^T$$

ماتریس قطری با مقادیر برابر با
eigenvalue ها روی قطر → نامنفی

$$k(x,y) = x A y = x M M^T y = (M^T x)^T (M^T y)$$

$$\phi(x) = M^T x \Rightarrow k(x,y) = \phi(x)^T \phi(y) = \langle \phi(x), \phi(y) \rangle \Rightarrow \text{Valid kernel}$$