DATE / / SUBJECT:

«به نام خدا»

امیررضا آذری — یادگیری ماشین

۹۹/۰۱/۰۸ — تمرین تئوری دوم

① کرنل معتبر

آ)

از تعریف کرنل داریم:

$K(x,y) = \phi(x)^T \phi(y) = \langle \phi(x), \phi(y) \rangle$

$\rightarrow K(x,z) = K_1(x,z) + K_2(x,z) = \langle \phi_1(x), \phi_1(z) \rangle + \langle \phi_2(x), \phi_2(z) \rangle = \left\langle \begin{bmatrix} \phi_1(x) \\ \phi_2(x) \end{bmatrix}, \begin{bmatrix} \phi_1(z) \\ \phi_2(z) \end{bmatrix} \right\rangle$

بنابراین $K(x,z)$ یک کرنل معتبر است.

ب)

$a \in R^+ \rightarrow K(x,z) = a K_1(x,z) = a \langle \phi_1(x), \phi_1(z) \rangle = \langle \sqrt{a}\, \phi_1(x), \sqrt{a}\, \phi_1(z) \rangle$

بنابراین این کرنل معتبر است.

ج)

$f: R^d \rightarrow R^p \rightarrow$ یک linear map است

$\rightarrow K(x,z) = K_3(f(x), f(z)) = \langle \phi_3(f(x)), \phi_3(f(z)) \rangle$

بنابراین این کرنل هم معتبر است.

CLASS MEMO

۲

آ) با تعدیل w_2 ؟

تغییری نخواهیم داشت و ثابت باقی می‌ماند. هنگام تعدیل w_2 ، مرز حاصل وابستگی کمتر و کمتری به x_2 پیدا می‌کند و با عمودی شدن این مرز داده‌های آموزش از یکدیگر جدا شده (با ارور صفر) و یک Vertical linear seperator خواهیم داشت.

ب) با تعدیل w_1 ؟

افزایش خواهد یافت. هنگام تعدیل w_1 ، مرز حاصل وابستگی کمتر و کمتری به x_1 پیدا می‌کند و بنابراین به حالت افقی نزدیک می‌شود. برای مقادیر بزرگ C ، خطای آموزش بیشتر شده و جداکننده افقی خطی خوبی نخواهد بود (بر روی داده آموزش)

ج) با تعدیل w_0 ؟

افزایش خواهد داشت. هنگام تعدیل w_0 ، در نهایت مرز به سمت origin می‌رود و به نوعی bias term برابر 0 می‌شود. مطابق شکل، با هیچ مرز خطی گذرنده از مبدأ نمی‌توان خطای صفر داشت. در این حالت خطا در بهترین حالت 1 خواهد بود.

د)

برای مقادیر بزرگ C ، هر دو w_1 و w_2 به 0 خواهند رفت. هنگامی که $w_1 = w_2 = 0$ ، log.probability لیبل ها ، یک مقدار متناهی به خود می‌گیرد. مثلا $w_0 = 0$ و برابر با $n\log(0.5)$.

در واقع $p(y=1|x,w) = p(y=0|x,w) = 0.5$. همین انتظار را هم داشتیم به دلیل اینکه تعداد عناصر در هر کلاس با هم برابر است و ما می‌خواهیم هر کدام را با احتمال برابر پیش‌بینی کنیم و $w_0 = 0$ ما را به $p(y=1|x,w) = 0.5$ می‌رساند.

ه)

برای مقادیر بزرگ C ، اشاره کردیم که هر دو w_2 و w_1 به صفر خواهند رفت. هنگامی که تعداد "+" ها بیشتر می‌شود و به‌نوعی از حالت بالانس خارج می‌شویم، ما خواهیم داشت که:

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

برای اینکه این امر اتفاق بیفتد، مقدار w_0 باید بیشتر از صفر باشد تا باعث $p(y=1|x,w) > 0.5$ خواهد شد.

DATE / / SUBJECT:

۳

الف)

$x, x' \in \mathbb{R}^n$

$$k(x,x') = \exp\left(-\frac{1}{2\sigma^2}\|x-x'\|^2\right) = \exp\left(-\frac{1}{2\sigma^2}\left(\|x\|^2 + \|x'\|^2 - 2\,x x'\right)\right)$$

$$= \exp\left(-\frac{1}{2\sigma^2}\|x\|^2 - \frac{1}{2\sigma^2}\|x'\|^2 + \frac{1}{\sigma^2}\,x x'\right)$$

$$= \underbrace{\exp\left(-\frac{1}{2\sigma^2}\|x\|^2\right)\exp\left(-\frac{1}{2\sigma^2}\|x'\|^2\right)}_{=A}\ \underbrace{\exp\left(+\frac{1}{\sigma^2}\,x\cdot x'\right)}_{=\sum_{i=0}^{\infty}\frac{(x\cdot x')^i}{i!\,\sigma^{2i}}}$$

$$= A \sum_{i=0}^{\infty} \sum_{d_1,\dots,d_n = i} \frac{1}{i!\,\sigma^{2i}} \frac{i!\prod_{j=1}^{n}(x_j \cdot x'_j)^{d_j}}{\prod_{j=1}^{n} d_j!}$$

$$= \sum_{i=0}^{\infty} \sum_{d_1,\dots,d_n = i} \frac{1}{i!\,\sigma^{2i}} \frac{i!\prod_{j=1}^{n} x_j^{d_j} \cdot x_j'^{d_j}}{\prod_{j=1}^{n} d_j!}\, A$$

در مرحله بعد، $i!$ ها ساده می‌شوند. همچنین $\sigma^{2i} = \sigma^i \sigma^i$.

$$= \sum_{i=0}^{\infty} \sum_{d_1,\dots,d_n = i} \frac{\prod_{j=1}^{n} x_j^{d_j}}{\sigma^i \prod_{j=1}^{n} d_j!} \exp\left(-\frac{1}{2\sigma^2}\|x\|^2\right) \frac{\prod_{j=1}^{n} x_j'^{d_j}}{\sigma^i \prod_{j=1}^{n} d_j!} \exp\left(-\frac{1}{2\sigma^2}\|x'\|^2\right)$$

$$= \sum_{i=0}^{\infty} \sum_{d_1,\dots,d_n = i} \phi_{d_1,\dots,d_n}(x)\, \phi_{d_1,\dots,d_n}(x')$$

بنابراین کرنل Gaussian را به صورت ضرب داخلی بردارها می‌توان نوشت.

ب)

A متقارن و مثبت نیمه معین است.

$$A = QDQ^T = \underbrace{Q\sqrt{D}}_{Q'}\sqrt{D}Q^T = Q'Q'^T$$

می‌دانیم این تجزیه مقادیر ویژه را قطری می‌کند و
هم‌چنین نامنفی است.

$$\to K(x,y) = x^TAy = x^TQ'Q'^Ty = (Q'^Tx)^T(Q'^Ty) = \langle Q'^Tx, Q'^Ty\rangle$$

به یک ضرب داخلی رسیدیم که در فضای تحت feature و $\Phi(x) = Q'^Tx$ هستیم. در نتیجه K یک
کرنل معتبر است. زیرا مثبت نیمه معین است. به دلیل اینکه تجزیه قطری دارای بردار ویژه‌های نامنفی است.

داریم:

$q \in R^p \longrightarrow$ show $q^T K q \geq 0$

$$\Rightarrow q^T K q = \sum_i \sum_j q_i (x_i^T A x_j) q_j = \sum_i \sum_j (q_i x_i)^T A (q_j x_j)$$

$$= \left(\sum_i q_i x_i\right)^T A \left(\sum_j q_j x_j\right)$$

$$= (q^T x)^T A (q^T x) = (x^T q)^T A (x^T q) \quad (I)$$

می دانیم A متقارن و مثبت نیمه معین است:

$$\xrightarrow{(I)} (x^T q)^T A (x^T q) = D^T A D \geq 0$$

در نتیجه داریم:

$$q^T K q = D^T A D \geq 0$$

بنابراین اثبات شد و این کل تمرین است.