

به نام خدا

عنوان:

تکلیف شماره هفتم درس یادگیری ماشین-SVM

استاد:

دکتر پدرام

دانشجو:

محمد علی مجتهد سلیمانی

شماره دانشجویی:

۴۰۳۳۹۰۴۵۰۴

تاریخ:

۱۴۰۳/۸/۲۰

سوال ۱.

سوال ۱

A) هر ترکیب خطی از توابع Kernel با فزاینده غیر منفی و مثبت

یک تابع Kernel خواهد بود پس جمع ۲ تابع Kernel نیز تابع

Kernel خواهد بود. ویژگی دیگر این است که بودن نیز رعایت شده است

$$\sum_{i=1}^n a_i u_i(x, z) = [u] = \sum_{i=1}^n a_i [u_i]$$

$$\text{for all } a \in \mathbb{R} \text{ and } a \geq 0 \quad a^T [u] a = \sum_{i=1}^n a_i a^T [u_i] a \geq 0$$

Kernel is a Positive-SemiDefinite

B) تفاوت دو تابع Kernel لزوماً یک تابع Kernel نمی شود

$$u_1 = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \quad u_2 = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$$

$$u = u_1 - u_2 = \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix} \quad \text{معادیر دیگره دارد}$$

$$\text{for } a \in \mathbb{R}^n \quad a^T [u] a \neq 0$$

*ادامه سوال در صفحه بعد قرار دارد.

C) از آنجایی که a غیر منفی است حاصل ضرب در a ضرب مثبت نیز فوکی kernel وجود

$$aa^T[k]a > 0$$

D) برعکس بالا از آنجایی که ضرب منفی است حاصل ضرب a منفی بود و kernel است.

$$aa^T[k]a \neq 0$$

E) $kernel$ متقارن است $k(x, x) = k(x, x)$ و باید $positive$ $semi$ $Definite$ باشد. اگر هم 2 باشد ضرب آن هم فوکی بود.

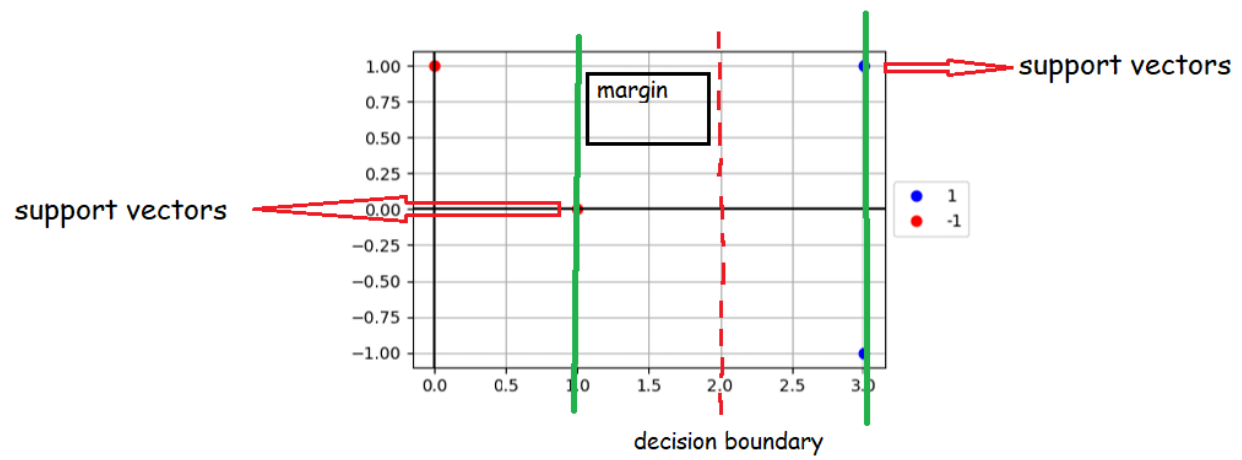
$$if \quad k_1 = \sum_i \mu_i u_i^T \quad if \quad k_2 = \sum_i \mu_i u_i^T$$

$$Product = k = \sum_i \mu_i^T u_i^T v_i \quad \rho = \mu^T$$

$$\omega = k \cdot v$$

$$So \Rightarrow aa^T[k]a > 0$$

سوال ۲.



*محاسبات در صفحه بعد قرار دارد.

محاسبات:

⑦

for linear SVM

$$w^T x + b = 0$$

$$\begin{cases} w^T x + b \geq 1 \\ w^T x + b \leq -1 \end{cases}$$

$$\begin{cases} w^T x + b \geq 1 \\ w^T x + b \leq -1 \end{cases}$$

class

$$\text{for } +1 = 3w_1 + w_2 + b \geq 1$$

class

$$\text{for } -1 = 3w_1 - w_2 + b \leq -1$$

$$\text{So } \Rightarrow \frac{1}{\|w\|^2} = \frac{1}{2} (w_1^2 + w_2^2)$$

بدون حل این معادله به صورت گوی

حل این معادله بدون داده هر افی توان دید

$$1 \leq \text{margin} \leq 3 \quad w = [1, 0], \quad b = -2$$

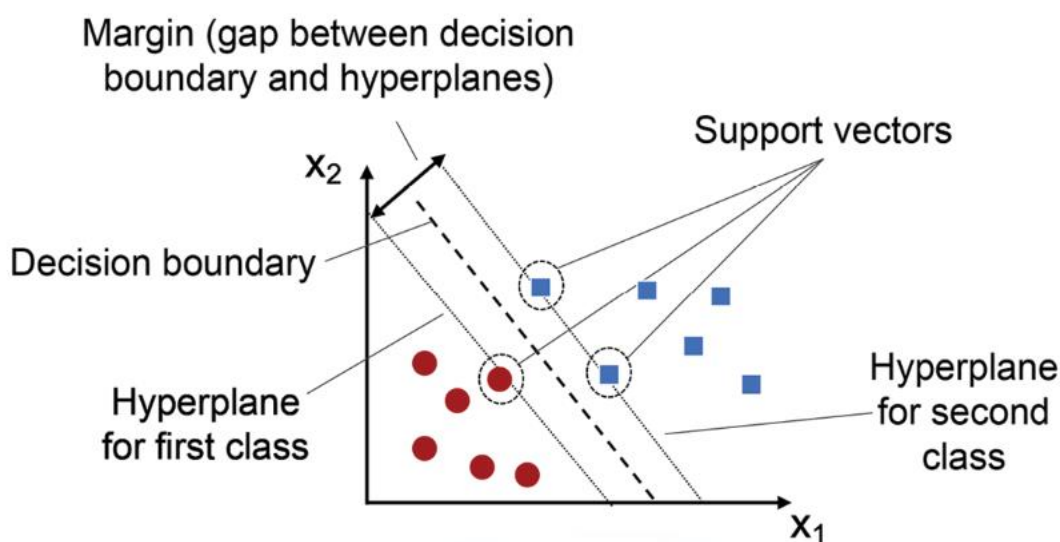
So $x_1 = 1$ for ex: for $(3, 1) : 3(1) + 0 - 2 \geq 1$

AYLAR

سوال ۳.

(الف)

درست است. بعد از آموزش SVM تنها چیزی که برای خط تصمیم گیری (decision boundary) مهم هست بردارهای پشتیبان (support vector) ها هستند و بقیه نقاط داده های آموزشی میتوانند بدون تاثیری در کارایی مدل حذف شوند. بخاطر اینکه فقط بردارهای پشتیبان هستند که بیشترین margin را میتوانند مشخص کنند. margin به فاصله بین بردارهای پشتیبان و خط تصمیم گیری گفته میشود و بردارهای پشتیبان به نزدیک ترین نقاط داده به خط تصمیم گیری گفته میشود.



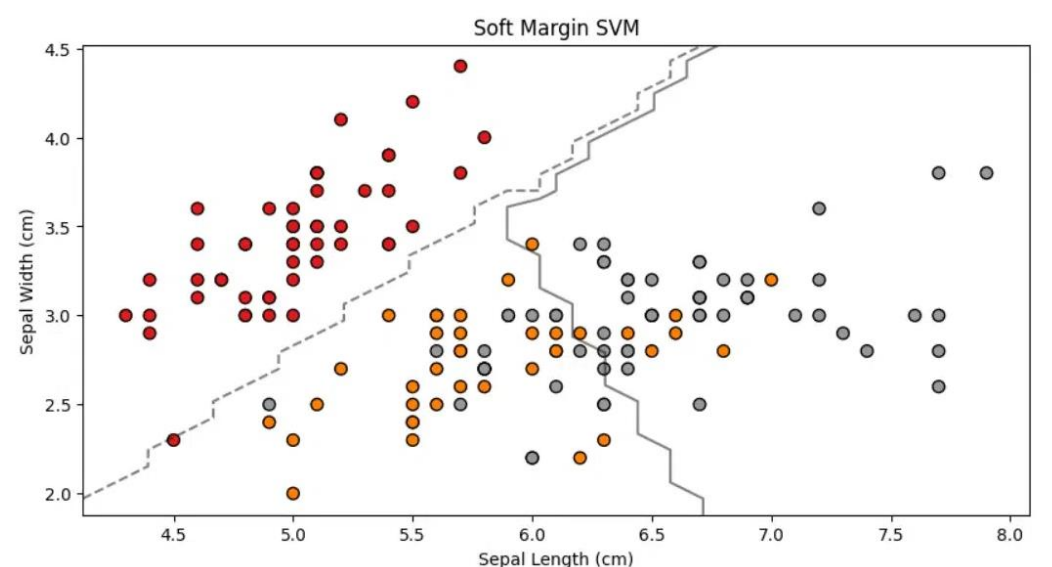
(ب)

درست است. تابع kernel در واقع نزدیکی بین ۲ داده ورودی را محاسبه میکند. این کار با محاسبه ضرب داخلی دو بردار (vector) در فضای ویژگی (feature space) انجام میشود. فضای ویژگی همان فضای مسئله ما هست اما در بعد های بالاتر یعنی بعد بیشتری به آن اضافه کردیم و مثلاً از ۲ بعد آن را به ۳ بعد بردیم.

سوال ۴.

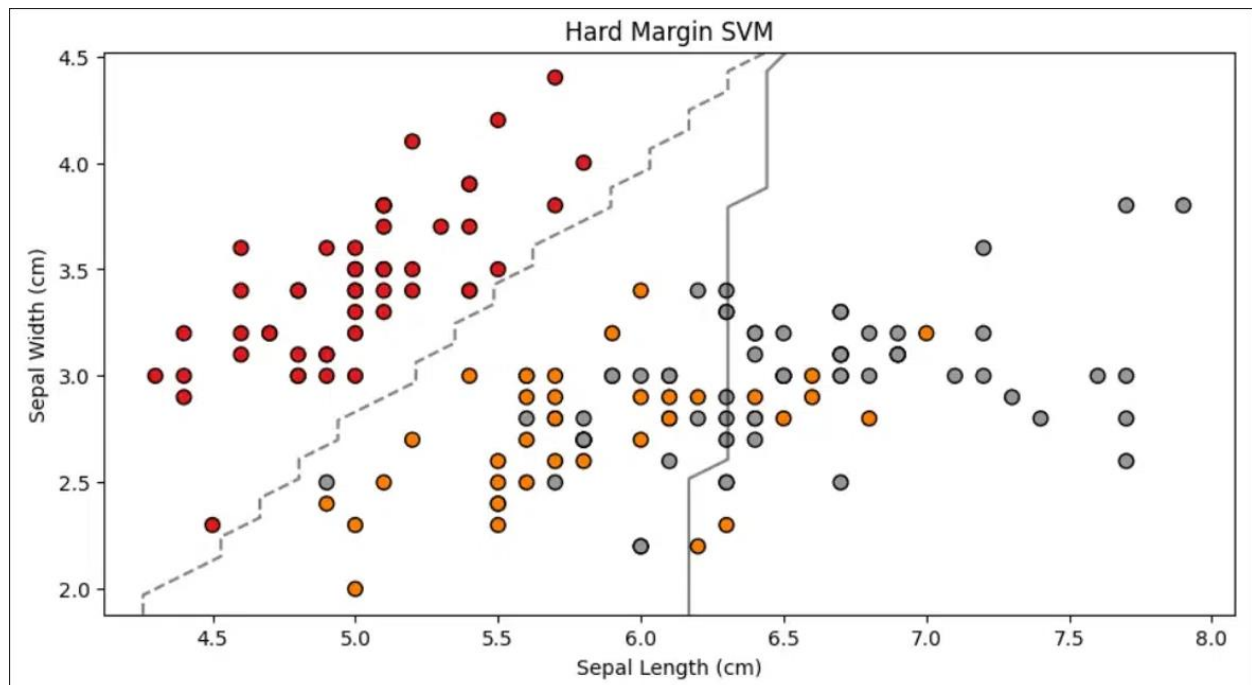
• Soft margin:

یک رویکردی است که اجازه میدهد که misclassification یا خطا در مجموعه آموزشی صورت بگیرد هنگامی که به دنبال مرز تصمیم گیری هستیم. این کار اجازه میدهد تا مدل ما overfit نباشد همچنین مرز تصمیم گیری ما نا پایدار نباشد. با اینکار مدل ما تعمیم پذیر خواهد بود نسبت به داده های جدیدی که آنها را ملاقات نکرده است. هم برای داده های خطی جدایی پذیر و غیر خطی جدا پذیر مناسب هستند. همچنین باعث یک trade off بین پیدا کردن بیشترین حد margin و کمترین مقدار خطا صورت بگیرد که اصطلاحاً low variance – high bias گفته میشود.



- Hard margin:

در طرف مقابل نسبت به **soft**، این یک رویکرد سخت گیر هست. که اجازه **misclassification** نمیدهد. به این معنی که داده ها باید به صورت خطی جدایی پذیر باشند. این رویکرد فقط به دنبال بیشینه کردن **margin** است و فقط از بردار های پشتیبان کمک میگیرد. در واقع این رویکرد هیچ تحمل پذیری در رابطه با **noise** ندارد و مستعد **overfit** شدن است.



سوال ۵.

در رویکرد آموزش، SVM ها به دنبال پیدا کردن یک مرز تصمیم گیری یا یک صفحه (hyperplane) برای جدا کردن داده های کلاس های مختلف هست که هدف آن بدست آوردن بیشترین margin است. در طرف دیگر شبکه های عصبی با کمک لایه های مختلف و تغییر مداوم پارامترهای آزاد مثل وزن ها یاد میگیرند.

SVM برای استفاده از dataset های کوچک تا متوسط مناسب هستند مخصوصا برای کلاس بندی های دودویی (binary). در طرف دیگر شبکه های عصبی بسیار نسبت منعطف هستند و میتوانند مدل های پیچیده تری به شمار بیان، و میتوانند روی مجموعه داده های بسیار بزرگ و بعد های بالاتر مورد استفاده قرار بگیرند. به همین دلیل SVM نیاز به منابع محاسباتی کمتری نسبت به شبکه های عصبی دارند. SVM ها هاپر پارامتر های کمتری نسبت به شبکه های عصبی دارند که فقط شامل C و پارامتر های kernel هست. همچنین انتخاب برای kernel هم بین مدل خطی و RBF محدود هستند. از طرف دیگر شبکه های عصبی با طیف گسترده ای از هاپر پارامترها و ... درگیر هستند.

SVM ها برای کلاس بندی متن، کلاس بندی دودویی و مجموعه داده های کوچک تا متوسط مناسب هستند.

در طرف دیگر شبکه های عصبی برای تشخیص الگو های پیچیده، مجموعه داده های بدون ساختار، کلاس بندی چند کلاسه مناسب هستند.

معمولا از SVM در پروژه های مانند spam email detection مورد استفاده قرار میگیرند اما در طرف دیگر شبکه های عصبی برای داده و پروژه های کار با عکس و ... مناسب هستند.