



دانشکده‌ی مهندسی کامپیوتر

یادگیری ماشین

بهار ۱۴۰۰

تمرین سری پنجم

مدرس: دکتر محمدحسین رهبان

زمان تحویل: ۱۹ اردیبهشت

سوال ۱ SVM

مساله‌ی ماشین بردار پشتیبان با حاشیه نرم^۱ را به شکل زیر در نظر بگیرید: (۲۰ نمره)

$$\min_{\mathbf{w}, b, \zeta} \quad \frac{1}{2} \|\mathbf{w}\|_2^2 + \frac{C}{2} \sum_{i=1}^N \zeta_i^2 \quad (1)$$

$$\text{subject to} \quad y^{(i)} (\mathbf{w}^T \mathbf{x}^{(i)} + b) \geq 1 - \zeta_i, \quad i \in \{1, 2, \dots, N\}$$

آ) برای $i \in \{1, 2, \dots, N\}$ نشان دهید شرط $\zeta_i \geq 0$ در جواب بهینه‌ی این مساله بی‌تاثیر است.

ب) این مساله را با استفاده از ضرایب لاگرانژ بازنویسی کرده و با مشتق‌گیری نسبت به پارامترهای \mathbf{w} , b و ζ عبارت لاگرانژی را بهینه کنید.

پ) با استفاده از بخش ب، فرم دوگان این مساله را بدست آورید.

ت) رفتار رابطه (۱) را در دو حالت $C \rightarrow 0$ و $C \rightarrow \infty$ تشریح کنید.

سوال ۲ Kernel

با فرض معتبر بودن هسته^۲ های k_1 و k_2 ، اعتبار هسته‌های زیر را بررسی کنید: (۱۰ نمره)

$$1. \quad k_3(x_1, x_2) = k_1(x_1, x_2) + k_2(x_1, x_2)$$

$$2. \quad k_4(x_1, x_2) = k_1(x_1, x_2) k_2(x_1, x_2)$$

$$3. \quad k_5(x_1, x_2) = e^{k_1(x_1, x_2)}$$

$$4. \quad k_6(x_1, x_2) = (1 - x_1^T x_2)^{-1}$$

¹Soft Margin Support Vector Machine (Soft Margin SVM)

²Kernel

سوال ۳ Properties of Kernels

می‌دانیم ویژگی‌های زیر برای هر هسته معتبر دلخواه مانند $f: \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ برقرار است:

$$\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d : f(\mathbf{x}, \mathbf{y}) = f(\mathbf{y}, \mathbf{x}) \quad \text{تقارنی}$$

$$\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^d : f(\mathbf{x} + \mathbf{z}, \mathbf{y}) = f(\mathbf{x}, \mathbf{y}) + f(\mathbf{z}, \mathbf{y}) \quad \text{افزایشی (برحسب اولین آرگومان)}$$

$$\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d, \alpha \in \mathbb{R}_{++} : f(\alpha \mathbf{x}, \mathbf{y}) = \alpha f(\mathbf{x}, \mathbf{y}) \quad \text{همگنی (برحسب اولین آرگومان)}$$

با این فرض که هسته $g: \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ هر سه ویژگی بالا را دارد، نشان دهید هسته $h: \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ که به شکل زیر تعریف می‌شود، معتبر است: (۱۰ نمره)

$$h(\mathbf{x}, \mathbf{y}) = \frac{1}{4} (g(\mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y}) - g(\mathbf{x} - \mathbf{y}, \mathbf{x} - \mathbf{y}))$$

سوال ۴ Kernels Over Sets

هسته‌ها نه تنها در فضای بردارها، بلکه در فضای مجموعه‌ها هم قابل استفاده هستند. برای مجموعه‌های دلخواه A و B ثابت کنید $k(A, B) = 2^{|A \cap B|}$ یک هسته معتبر در فضای مجموعه‌هاست. (۱۰ نمره)

سوال ۵ (عملی) فایل *Notebook*ی که در اختیار شما قرار داده شده را کامل کنید. در این تمرین قصد داریم با پیاده‌سازی الگوریتم *SVM* آشنا شده و به دسته‌بندی داده‌هایی که در فایل *Heart_Disease_Dataset.csv* قرار دارند بپردازیم. توصیفی از این مجموعه داده در فایل *Dataset_Description.pdf* آمده است. برای دریافت نمره کامل بخش *MySVM*، باید دقت بهترین هسته پیاده‌سازی شده روی داده‌های آزمون^۱، دست‌کم ۹۰ درصد باشد. برای دقت‌های بالاتر روی این داده‌ها، در صورتی که چرایی بهبود مدل را شرح دهید، نمره امتیازی متناسب با افزایش دقت در نظر گرفته می‌شود. (۴۰ نمره + ۵ نمره امتیازی)

پاینده باشید

¹Test Data