



دانشکده مهندسی سامانه‌های هوشمند

گروه علوم داده

یادگیری ماشین

تمرین پنجم

استاد درس

دکتر سامان هراتی‌زاده

زمان تحویل: ۱۴۰۳/۱۰/۰۸

یادگیری ماشین – تمرین «»5



ددلاین: 1403/10/08

دستیاران آموزشی
[سجاد دشتی](#)

دکتر سامان هراتی زاده
دانشگاه تهران - دانشکده سامانه‌های هوشمند
نیم‌سال اول ۱۴۰۳-۱۴۰۴

مقدمه و نکات

- لطفاً پاسخ سوالات تحلیلی را با کیفیت مناسب در گزارش قرار دهید.
- دادگان مورد نیاز را از کتابخانه scikit-learn بدست آورید.
- در این تمرین شما محدودیتی در استفاده از پکیج‌های مورد نیاز ندارید.
- لطفاً گزارش را در قالب مشخص شده نوشته و با فرمت pdf تحویل دهید.
- تمامی اجزا کد و نتایج می‌بایست توسط مصححین، عیناً تکرار پذیر باشند (حتی برای بخش بندی دادگان). می‌توانید از random_state برای تکرار پذیری استفاده کنید.
- تحلیل نتایج حائز اهمیت است. لذا در تمامی قسمت‌ها نتایج به‌دست‌آمده از تمرین باید حتماً در گزارش درج و تحلیل شوند.
- در حل این تمرین، شما مجاز به استفاده از مدل‌های زبانی بزرگ (LLM) برای کمک به نوشتن کد یا حل مسائل هستید. با این حال، شما باید به‌طور کامل به تمامی کدی که تحویل می‌دهید تسلط داشته باشید و قادر به توضیح عملکرد و نیز تغییر کد باشید. استفاده از ابزارها و مدل‌های کمکی به این معنا نیست که بتوانید بدون درک کافی از کد آن را ارائه دهید؛ هدف این است که دانش و درک عمیقی از مفاهیم و راه‌حل‌ها داشته باشید.
- توضیح مختصری از نحوه عملکرد اجزای کلیدی کد ضروری است.
- برای سوالات خود می‌توانید از طریق ایمیل s.dashti.k@gmail.com و نیز گروه تلگرامی اقدام بفرمایید.

یادگیری ماشین – تمرین «»5



ددلاین: 1403/10/08

دستیاران آموزشی
سجاد دشتی

دکتر سامان هراتی زاده
دانشگاه تهران - دانشکده سامانه‌های هوشمند
نیمسال اول ۱۴۰۳-۱۴۰۴

سوالات تحلیلی:

۱. نگاشت از فضای ورودی به فضای با ابعاد بالاتر به صورت زیر است:

$$x \in R \rightarrow y \equiv \varphi(x) \in R^{2k+1}$$

$$\varphi(x) = \left[\frac{1}{\sqrt{2}}, \cos x, \cos 2x, \dots, \cos kx, \sin x, \sin 2x, \dots, \sin kx \right]^T$$

نشان دهید که کرنل ضرب داخلی مربوطه به صورت زیر است:

$$k(x_i, x_j) = y_i^T y_j = \frac{\sin((k + 0.5)(x_i - x_j))}{2 \sin(\frac{x_i - x_j}{2})}$$

یادگیری ماشین – تمرین «5»



ددلاین: 1403/10/08

دستیاران آموزشی
سجاد دشتی

دکتر سامان هراتی‌زاده
دانشگاه تهران - دانشکده سامانه‌های هوشمند
نیم‌سال اول ۱۴۰۳-۱۴۰۴

سوالات پیاده سازی:

۲. کرنل Radial Basis Function یا همان RBF یکی از پرکاربردترین توابع کرنل در ماشین‌های بردار پشتیبان (SVM) است. این تابع به‌ویژه در مسائل غیرخطی مؤثر است؛ زیرا داده‌های ورودی را به فضایی با ابعاد بالاتر نگاشت می‌کند و به SVM امکان می‌دهد ابرصفحه بهینه‌ای را برای جداسازی کلاس‌های مختلف بیابد.

از نظر ریاضی، کرنل RBF بین دو نمونه به‌صورت زیر تعریف می‌شود:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2)$$

$$\gamma > 0$$

الف) ابتدا مختصراً توضیحی در مورد هر کدام از کرنل‌های Linear و Polynomial دهید و بگویید هر کدام برای چه نوع داده یا مساله ای بهتر عمل می‌کند.

ب) با استفاده از دادگان fetch_covtype و ویژگی‌هایی که در پایین داده شده اند از طریق کرنل rbf برای مقادیر مختلف دادگان آموزشی که در لیست زیر داده شده اند، مدل‌هایی را آموزش دهید و برای هر مدل، زمان اجرا را ذخیره کنید. سپس نمودار زمان آموزش بر اساس تعداد داده‌های آموزش را رسم کنید. (از روش one vs one استفاده شود که این کتابخانه بطور پیشفرض انجام می‌دهد)

ویژگی‌ها: Elevation , Aspect , Slope , Horizontal_Distance_To_Hydrology

تعداد دادگان آموزش:

100, 500, 1000, 2500, 5000, 7500, 10000, 15000, 20000, 25000, 30000, 35000,
40000, 45000, 50000

یادگیری ماشین – تمرین «5»



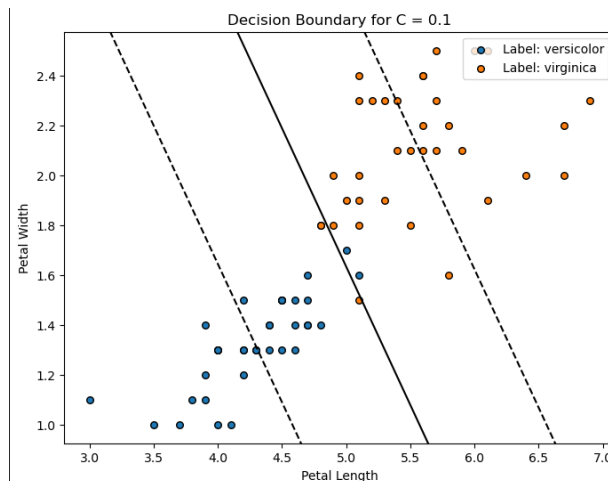
ددلاین: 1403/10/08

دستیاران آموزشی
سجاد دشتی

دکتر سامان هراتی زاده
دانشگاه تهران - دانشکده سامانه‌های هوشمند
نیم‌سال اول ۱۴۰۳-۱۴۰۴

ج) توضیح دهید که پارامتر C چه نقشی در الگوریتم SVM دارد. روی دیتاست iris، با استفاده از کرنل Linear و دادگان مربوط به دو کلاس versicolor و virginica بطوریکه ۳۰ درصد دادگان مربوط به تست باشند و از دو ویژگی Petal Length و Petal Width استفاده شود، برای مقادیر مختلف C شامل ۱۰۰، ۱۰، ۱، ۰/۱، ۰/۰۱، ۰/۰۰۱ مدل هایی را آموزش دهید و مراحل زیر را انجام دهید:

- دقت مدل‌ها روی داده‌های Train و Test را گزارش کنید.
- به ازای هر مدل، یک تصویر شامل دادگان آموزشی، مرز تصمیم و صفحه های حاشیه ای رسم نمایید. (دادگان آموزشی هر کلاس رنگ متفاوت داشته باشند)
- نمودار روند Accuracy آموزش و تست بر اساس C را گزارش کنید.



(نمونه ای از تصویری که باید گزارش شود)

یادگیری ماشین – تمرین «»5



ددلاین: 1403/10/08

دستیاران آموزشی
سجاد دشتی

دکتر سامان هراتی‌زاده
دانشگاه تهران - دانشکده سامانه‌های هوشمند
نیم‌سال اول ۱۴۰۳-۱۴۰۴

(د) با استفاده از دادگانی که در زیر توضیح داده شده است، برای کرنل rbf به ازای مقادیر gamma که در پایین داده شده اند، مدل آموزش دهید.

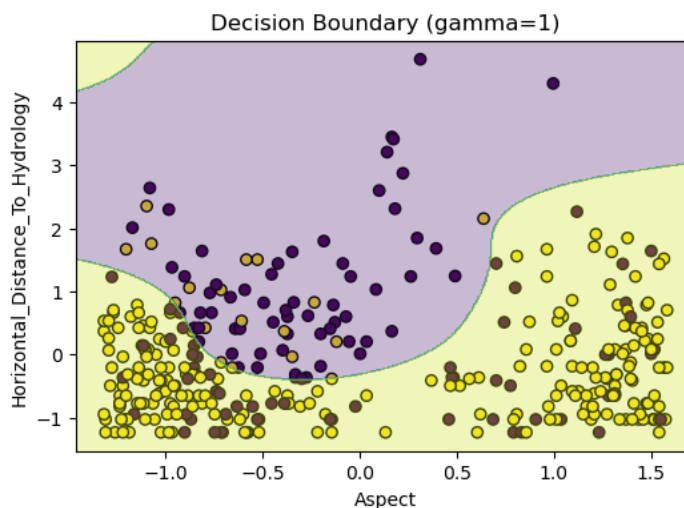
[0.1, 0.2, 0.5, 1, 5, 10, 20, 40, 60, 80, 100]

دادگان:

مجموعه داده مورد استفاده fetch_covtype است و می بایست ویژگی های Aspect و Horizontal_Distance_To_Hydrology و کلاس های پنجم و ششم مورد استفاده قرار گیرند. ۵۰۰ داده را از این دادگان را انتخاب کرده و ۳۰ درصد از این دادگان انتخاب شده برای تست و ۷۰ درصد آموزش باشد. حتما نرمال سازی انجام شود.

سپس ناحیه هر کلاس از نظر مدل نسبت به ویژگی های نرمال استفاده شده را همراه با دادگان آموزش به ازای gamma شامل ۰/۱ و ۱ و ۱۰ و ۴۰ رسم کنید.

همچنین نمودار دقت تست و آموزش به ازای gamma را رسم کنید.



(نمونه ای از تصویری که باید گزارش شود)

یادگیری ماشین – تمرین «»5



ددلاین: 1403/10/08

دستیاران آموزشی
[سجاد دشتی](#)

دکتر سامان هراتی زاده
دانشگاه تهران - دانشکده سامانه‌های هوشمند
نیم‌سال اول ۱۴۰۳-۱۴۰۴

هـ) با استفاده از دادگان بخش قبل با این تفاوت که از ۵۰۰ داده انتخاب شده، ۵۰ درصد برای آموزش و ۵۰ درصد برای تست باشند، برای کرنل چند جمله ای به ازای مقادیر زیر برای درجه چند جمله ای، مدل آموزش دهید.

درجه: ۱، ۲، ۳، ۴، ۶، ۸، ۱۰، ۱۲، ۱۴، ۱۶

سپس ناحیه هر کلاس از نظر مدل نسبت به ویژگی های نرمال استفاده شده را همراه با دادگان آموزش به ازای درجه های ۲، ۸ و ۱۶ رسم کنید.

همچنین نمودار دقت تست و آموزش به ازای درجه را رسم کنید.

یادگیری ماشین – تمرین «»5



ددلاین: 1403/10/08

دستیاران آموزشی
[سجاد دشتی](#)

دکتر سامان هراتی زاده
دانشگاه تهران - دانشکده سامانه‌های هوشمند
نیم‌سال اول ۱۴۰۳-۱۴۰۴

و) روی دادگان بخش د، برای حالت های زیر، ناحیه کلاسهای مختلف (از نظر مدل) نسبت به ویژگی های استفاده شده را به همراه دادگان آموزش رسم کنید.

- SVM with Linear Kernel
- SVM with RBF Kernel $\gamma=5$
- SVM with Polynomial Kernel $d=10$

برای همه مدل ها، دقت روی دادگان آموزش و تست را روی یک نمودار میله ای نشان دهید و مدلها را با هم مقایسه کنید.

توضیح دهید کرنل هایی که استفاده شد، دادگان را به فضای چند بعدی می برند و اثر این موضوع چیست؟