

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:

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

$$\varphi(x_i)^T \varphi(x_j) = \left(\frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} \right) + \sum_{n=1}^k (\cos(nx_i)\cos(nx_j) + \sin(nx_i)\sin(nx_j))$$

$$\cos(nx_i)\cos(nx_j) + \sin(nx_i)\sin(nx_j) = \cos(n(x_i - x_j))$$

$$\varphi(x_i)^T \varphi(x_j) = \frac{1}{2} + \sum_{n=1}^k \cos(n(x_i - x_j))$$

$$\sum_{n=1}^k \cos(n\theta) = \frac{\sin((k + 0.5)\theta)}{2\sin(\theta/2)}$$

$$\sum_{n=1}^k \cos(n(x_i - x_j)) = \frac{\sin((k + 0.5)(x_i - x_j))}{2\sin((x_i - x_j)/2)}$$

$$\varphi(x_i)^T \varphi(x_j) = \frac{1}{2} + \frac{\sin((k + 0.5)(x_i - x_j))}{2\sin((x_i - x_j)/2)}$$

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

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:

$$k(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$$

$$\gamma > 0$$

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

کرنل خطی

کرنل خطی به شکل ساده‌ای داده‌ها را به یک فضای با ابعاد بالاتر نگاشت می‌کند و به صورت زیر تعریف می‌شود:

$$k(x_i, x_j) = x_i^T x_j$$

کاربردها:

- مسائل با تعداد ویژگی‌های بالا: کرنل خطی برای داده‌هایی که تعداد ویژگی‌هایشان بسیار بیشتر از تعداد نمونه‌ها است، بسیار مناسب است، زیرا پیچیدگی محاسباتی کم‌تری دارد.
- داده‌های قابل تفکیک خطی: اگر داده‌ها به صورت خطی قابل تفکیک باشند، کرنل خطی به خوبی عمل می‌کند.
- نرم‌های خطی: در مسائلی که تصمیم‌گیری بر اساس ترکیب خطی ویژگی‌ها انجام می‌شود، کرنل خطی مناسب است.

مثال‌ها:

- **طبقه‌بندی متن**: در مسائل طبقه‌بندی متن، ویژگی‌ها (مثل تعداد کلمات) بسیار زیاد هستند و کرنل خطی به خوبی عمل می‌کند.
- **تحلیل مالی**: در مسائل مرتبط با تحلیل مالی که ویژگی‌ها خطی هستند و داده‌ها نیز به صورت خطی قابل تفکیک می‌باشند.

کرنل Polynomial

کرنل چندجمله‌ای به صورت زیر تعریف می‌شود:

$$k(x_i, x_j) = (\gamma x_i^T x_j + r)^d$$

که در آن:

- γ پارامتر مقیاس است.
- r جابجایی است.
- d درجه چندجمله‌ای است.

کاربردها:

- **مسائل غیرخطی**: کرنل چندجمله‌ای برای داده‌هایی که به صورت غیرخطی قابل تفکیک هستند، مناسب است.
- **داده‌های پیچیده**: در مسائلی که داده‌ها الگوهای پیچیده و غیرخطی دارند، این کرنل مفید است.
- **ارتباطات درجه بالا**: در مسائلی که روابط پیچیده بین ویژگی‌ها وجود دارد، این کرنل قادر به مدل‌سازی این روابط است.

مثال‌ها:

- **طبقه‌بندی تصویر**: در مسائل طبقه‌بندی تصویر که الگوهای پیچیده‌ای در داده‌ها وجود دارد، کرنل چندجمله‌ای کاربرد دارد.

- **تحلیل ژنومی**: در مسائل مرتبط با تحلیل داده‌های زیستی که الگوهای پیچیده و غیرخطی بین داده‌ها وجود دارد.

مقایسه کلی

- **پیچیدگی محاسباتی**: کرنل خطی ساده‌تر و سریع‌تر است، در حالی که کرنل چندجمله‌ای پیچیدگی بیشتری دارد.
- **توانایی مدل‌سازی**: کرنل چندجمله‌ای توانایی مدل‌سازی روابط پیچیده‌تر را دارد.

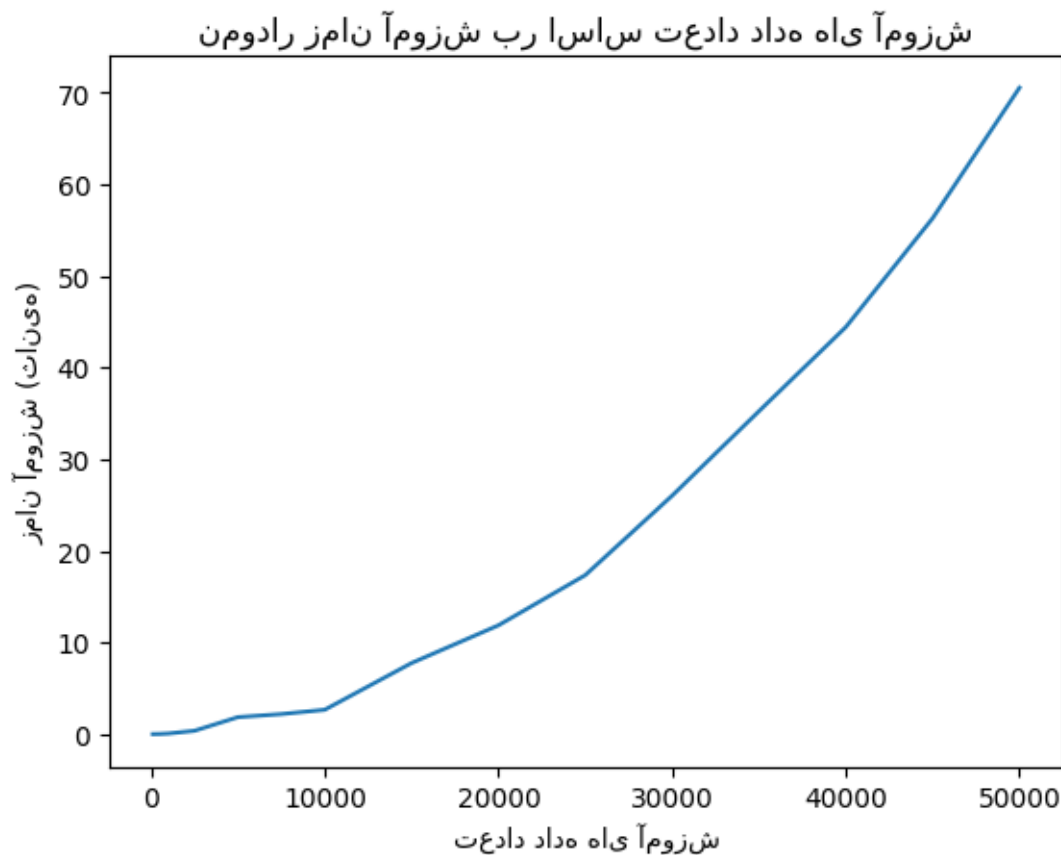
- کاربرد: کرنل خطی بیشتر برای مسائل ساده‌تر و داده‌های قابل تفکیک خطی مناسب است، در حالی که کرنل چندجمله‌ای برای داده‌های پیچیده و غیرخطی کاربرد دارد.

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

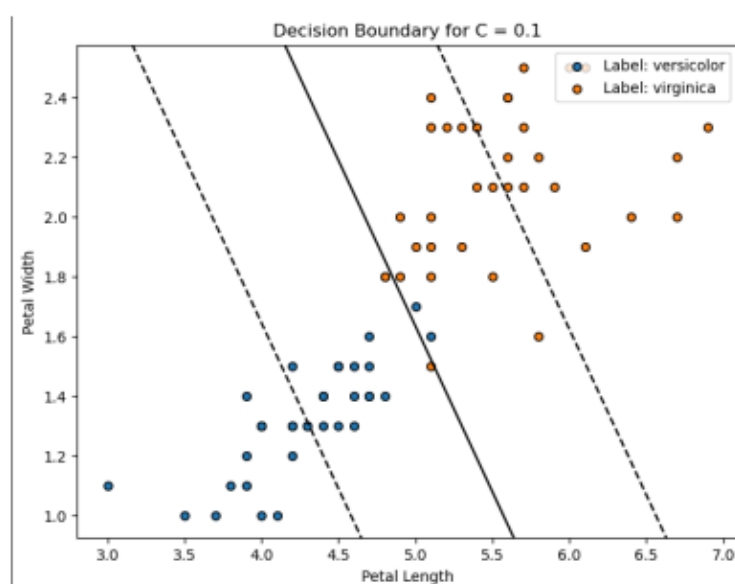
تعداد دادگان آموزش:

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



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 شامل مدل هایی را آموزش دهید و مراحل زیر را انجام دهید:

- دقت مدل ها روی داده های 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)

پارامتر C در الگوریتم SVM یک پارامتر تنظیم کننده است که میزان جریمه خطاهای طبقه بندی را کنترل می کند. این پارامتر تعادلی بین دقت مدل و پیچیدگی آن ایجاد می کند. به عبارت دیگر، پارامتر C بین بیش برآزش (overfitting) و کم برآزش (underfitting) توازن برقرار می کند.

نقش پارامتر C در جزئیات:

- **مقادیر کوچک C :** با انتخاب مقادیر کوچک برای C ، به مدل اجازه داده می شود که خطاهای طبقه بندی بیشتری را تحمل کند. این کار باعث می شود تا مرز تصمیم گیری هموارتر باشد و مدل کمتر به داده های آموزشی حساس باشد. در نتیجه، خطر بیش برآزش کاهش می یابد، اما ممکن است مدل دقت کمتری داشته باشد.

- **مقادیر بزرگ C:** با انتخاب مقادیر بزرگ برای C، مدل جریمه بیشتری برای خطاهای طبقه‌بندی در نظر می‌گیرد. این کار باعث می‌شود تا مرز تصمیم‌گیری پیچیده‌تر شود و مدل بیشتر به داده‌های آموزشی حساس باشد. در نتیجه، دقت مدل افزایش می‌یابد، اما خطر بیش‌برازش افزایش می‌یابد.

انتخاب مقدار مناسب برای C:

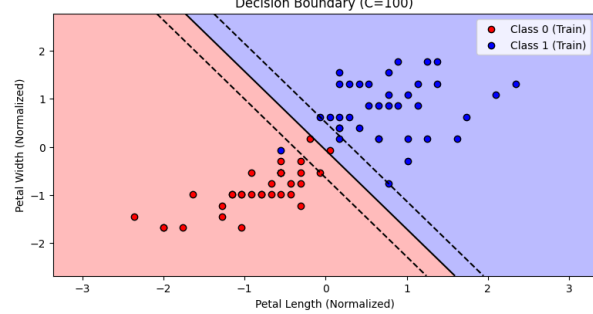
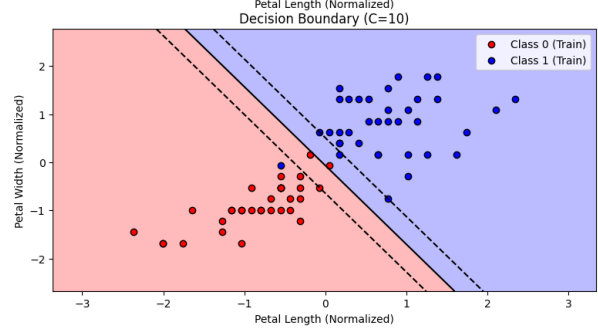
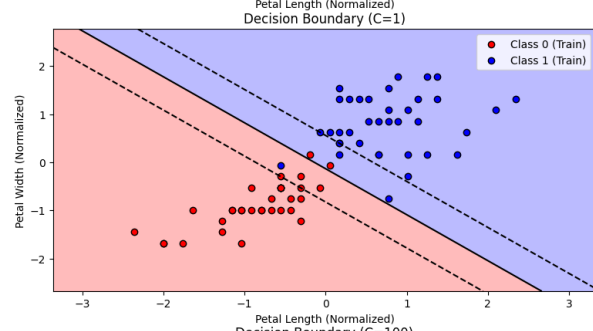
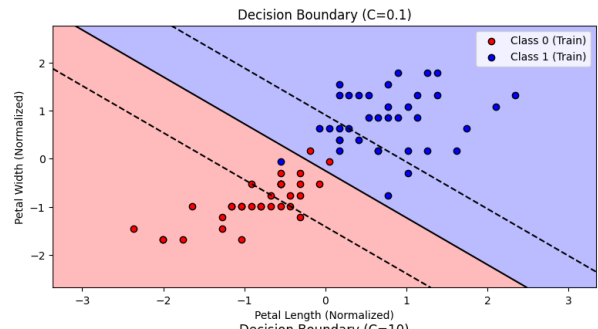
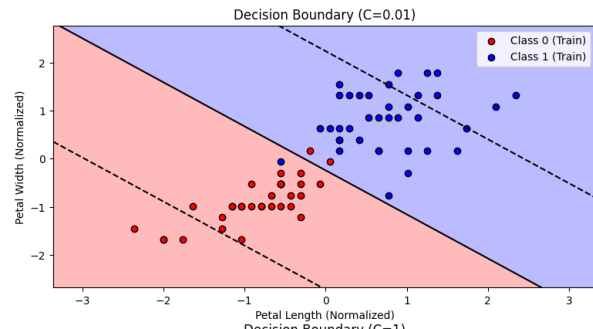
انتخاب مقدار مناسب برای C به داده‌های آموزشی و نوع مسئله بستگی دارد. به طور کلی، برای داده‌هایی که پیچیدگی کمتری دارند، مقادیر کوچک‌تر C مناسب است. برای داده‌هایی که پیچیدگی بیشتری دارند، مقادیر بزرگ‌تر C مناسب است.

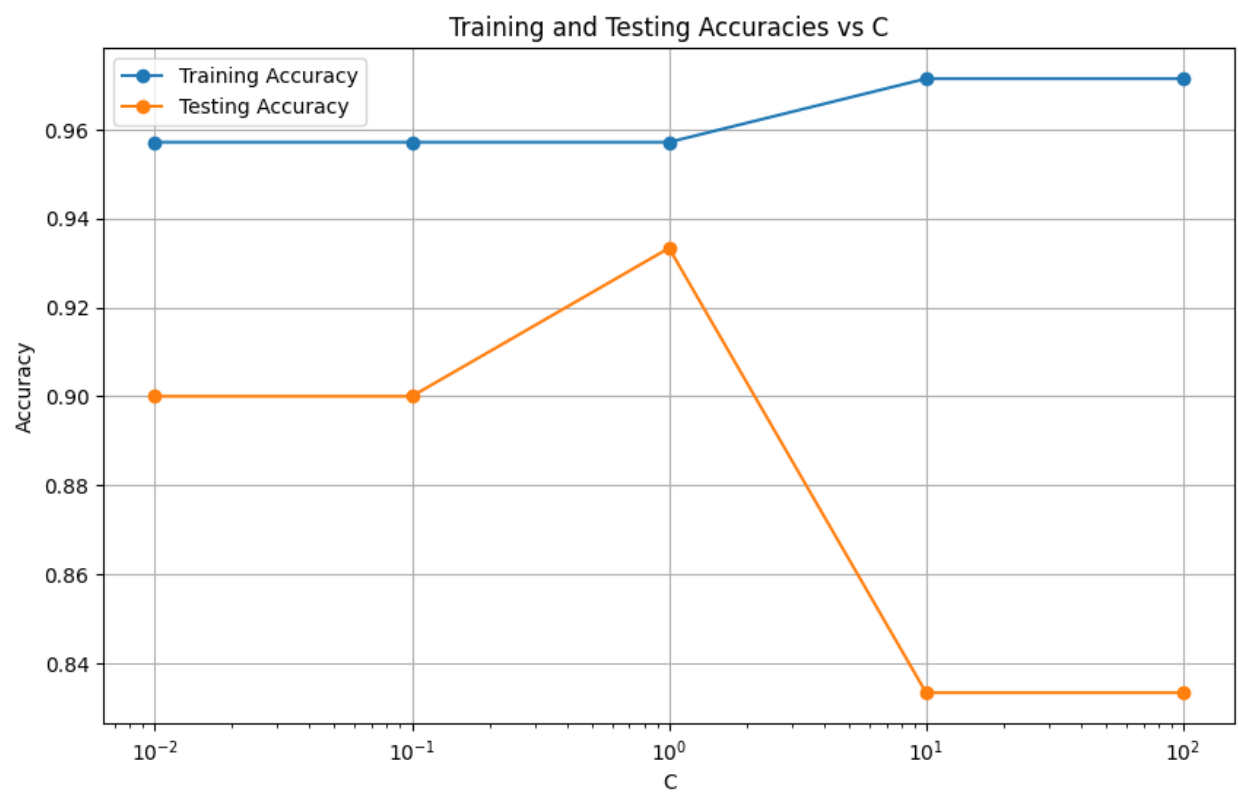
روش‌های انتخاب C:

- **اعتبار سنجی متقاطع (Cross-validation):** این روش یکی از رایج‌ترین روش‌ها برای انتخاب مقدار مناسب C است. در این روش، داده‌های آموزشی به چندین بخش تقسیم می‌شوند و مدل برای مقادیر مختلف C آموزش داده می‌شود. سپس، دقت مدل برای هر مقدار C روی داده‌های اعتبارسنجی (بخشی از داده‌های آموزشی که برای آموزش مدل استفاده نشده است) ارزیابی می‌شود. در نهایت، مقداری از C که بهترین دقت را روی داده‌های اعتبارسنجی داشته باشد، انتخاب می‌شود.
- **Grid Search:** این روش شامل امتحان کردن مجموعه‌ای از مقادیر مختلف برای C و ارزیابی دقت مدل برای هر مقدار است. در نهایت، مقداری از C که بهترین دقت را داشته باشد، انتخاب می‌شود.

نتیجه‌گیری:

پارامتر C در الگوریتم SVM نقش مهمی در تنظیم تعادل بین دقت و پیچیدگی مدل دارد. انتخاب مقدار مناسب برای C به داده‌های آموزشی و نوع مسئله بستگی دارد و باید با استفاده از روش‌های مناسب مانند اعتبارسنجی متقاطع یا Grid Search انجام شود.





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.

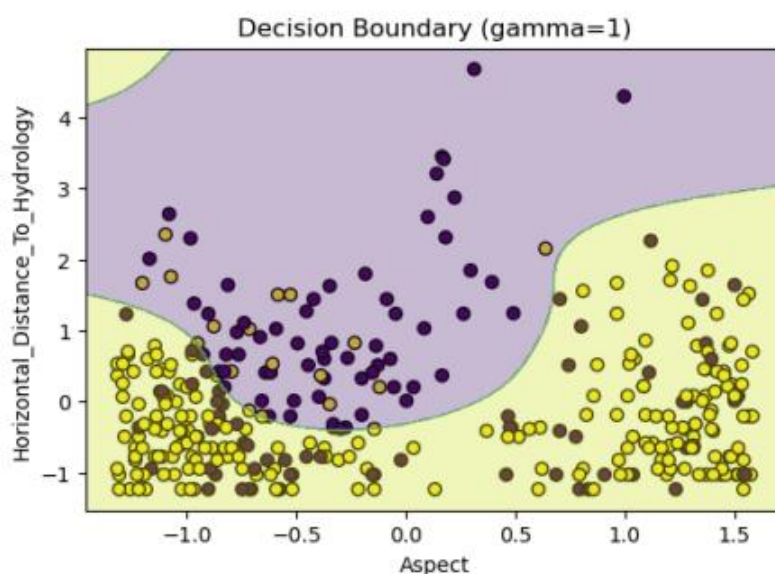
[0.1, 0.2, 0.5, 1, 5, 10, 20, 40, 60, 80, 100]

دادگان:

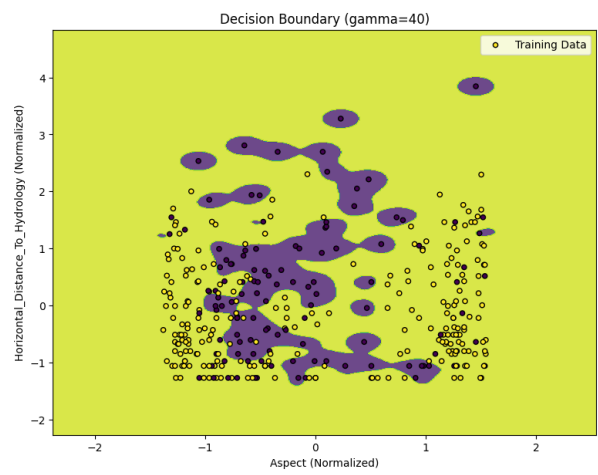
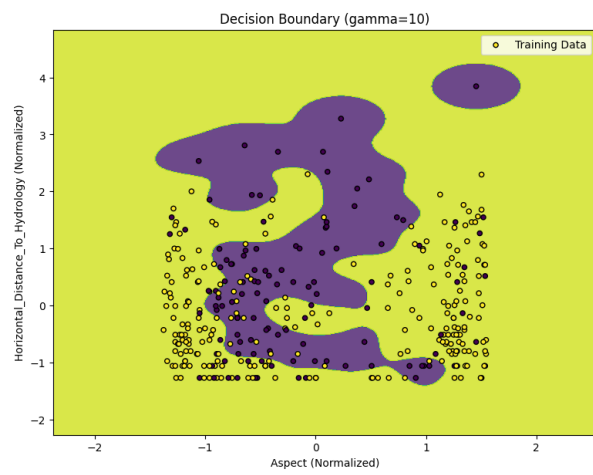
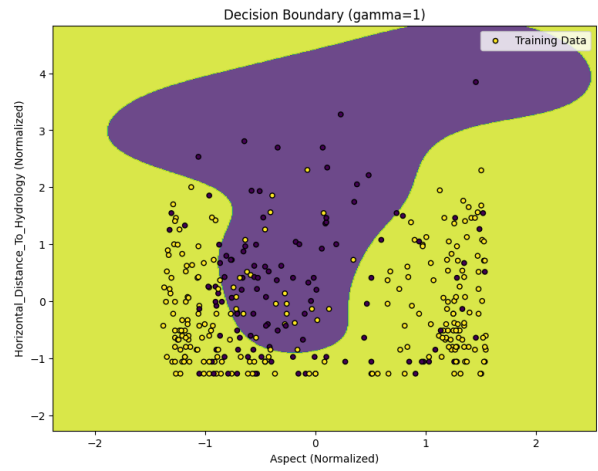
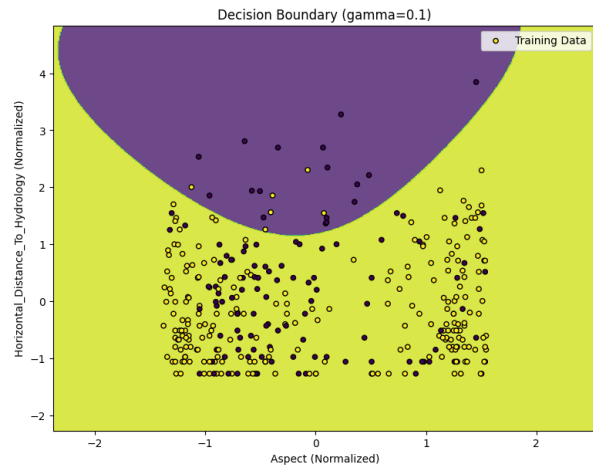
مجموعه داده مورد استفاده fetch_covtype است و می بایست ویژگی های Aspect و Horizontal_Distance_To_Hydrology و کلاس های پنجم و ششم مورد استفاده قرار گیرند. ۵۰۰ داده را از این دادگان را انتخاب کرده و ۳۰ درصد از این دادگان انتخاب شده برای تست و ۷۰ درصد آموزش باشد. حتما نرمال سازی انجام شود.

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)

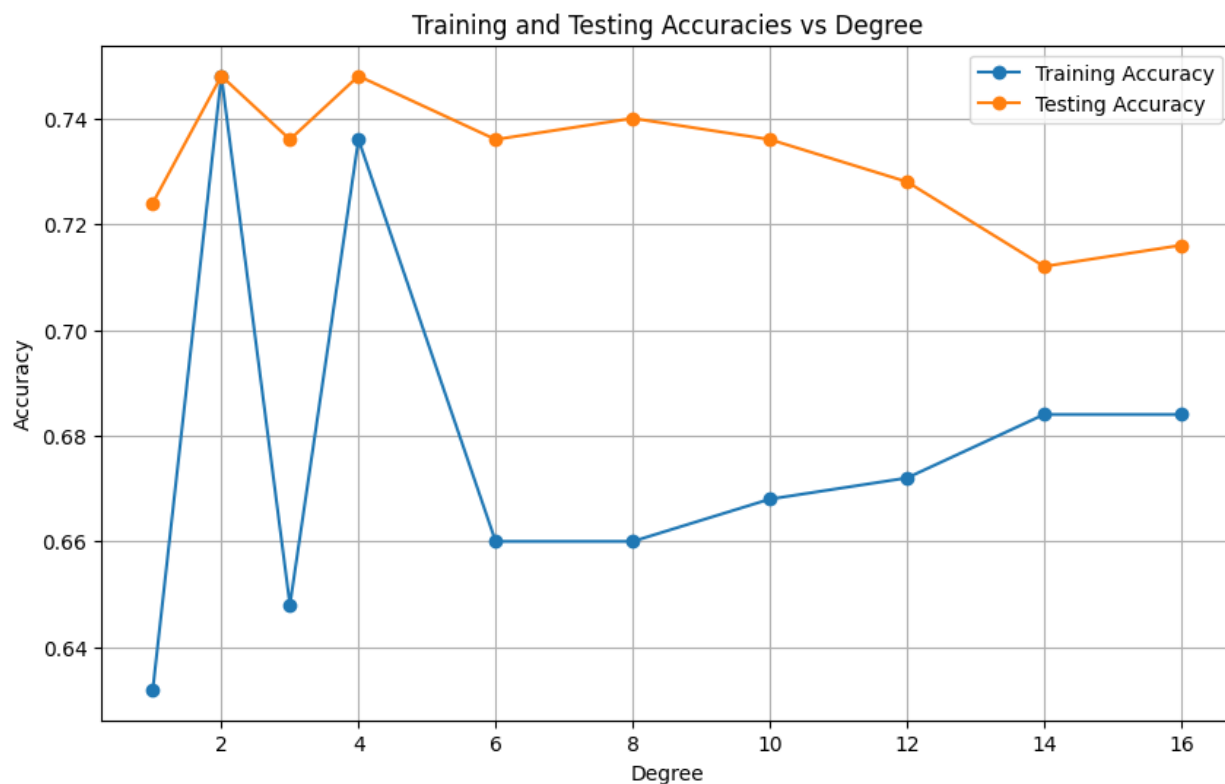


ه) با استفاده از دادگان بخش قبل با این تفاوت که از ۵۰۰ داده انتخاب شده، ۵۰ درصد برای آموزش و ۵۰ درصد برای تست باشند، برای کرنل چند جمله ای به ازای مقادیر زیر برای درجه چند جمله ای، مدل آموزش دهید.

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.

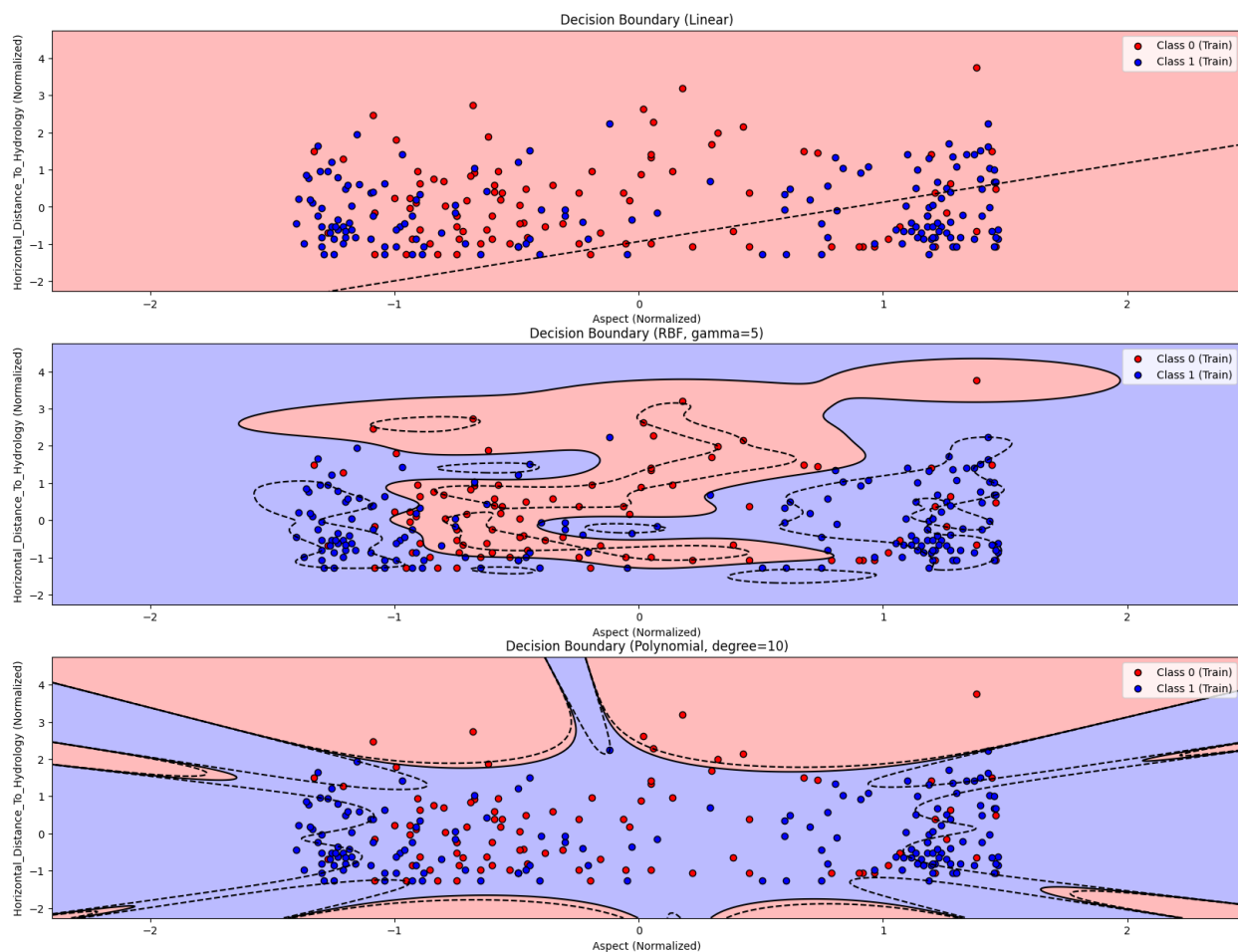


و) روی دادگان بخش د، برای حالت های زیر، ناحیه کلاسه های مختلف (از نظر مدل) نسبت به ویژگی های استفاده شده را به همراه دادگان آموزش رسم کنید.

- 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؟



چگونه کرنل‌ها داده‌ها را به فضای چند بعدی می‌برند؟

به جای محاسبه صریح تبدیل داده‌ها به فضای با ابعاد بالاتر، کرنل‌ها از یک تابع شباهت بین جفت نقاط در فضای اصلی استفاده می‌کنند. این تابع شباهت، حاصل ضرب داخلی بین دو نقطه را در فضای با ابعاد بالاتر محاسبه می‌کند، بدون اینکه نیازی به محاسبه صریح تبدیل داده‌ها باشد. به عبارت دیگر، کرنل‌ها به صورت غیرمستقیم داده‌ها را به فضای چند بعدی می‌برند.

اثر انتقال داده‌ها به فضای چند بعدی چیست؟

با انتقال داده‌ها به فضای با ابعاد بالاتر، امکان تفکیک خطی داده‌هایی که در فضای اصلی قابل تفکیک خطی نیستند، فراهم می‌شود. این امر به SVM اجازه می‌دهد تا مرزهای تصمیم‌گیری پیچیده‌تری را برای طبقه‌بندی داده‌ها ایجاد کند.

مثال:

فرض کنید داده‌هایی داریم که در فضای دو بعدی قابل تفکیک خطی نیستند. با استفاده از یک کرنل مناسب، می‌توانیم این داده‌ها را به فضای سه بعدی منتقل کنیم، به طوری که در فضای سه بعدی قابل تفکیک خطی باشند. در این حالت، SVM می‌تواند یک صفحه تفکیک خطی در فضای سه بعدی ایجاد کند که در فضای اصلی به صورت یک مرز تصمیم‌گیری غیرخطی نمایش داده می‌شود.

انواع کرنل‌ها:

- **کرنل خطی:** این کرنل ساده‌ترین نوع کرنل است و برای داده‌هایی که به صورت خطی قابل تفکیک هستند، مناسب است.
- **کرنل چندجمله‌ای:** این کرنل برای داده‌هایی که به صورت غیرخطی قابل تفکیک هستند، مناسب است و می‌تواند مرزهای تصمیم‌گیری پیچیده‌تری را ایجاد کند.
- **کرنل RBF:** این کرنل نیز برای داده‌های غیرخطی مناسب است و می‌تواند مرزهای تصمیم‌گیری بسیار پیچیده‌ای را ایجاد کند.

نتیجه‌گیری:

کرنل‌ها نقش مهمی در SVM دارند و به این الگوریتم اجازه می‌دهند تا داده‌های غیرخطی را به صورت خطی تفکیک کند. این امر باعث می‌شود تا SVM بتواند مرزهای تصمیم‌گیری پیچیده‌تری را برای طبقه‌بندی داده‌ها ایجاد کند و دقت بالاتری را در مسائل طبقه‌بندی بدست آورد.