University of Tehran
School of Electrical and Computer Engineering

MACHINE LEARNING

Assignment 4

Sasan Keshavarz
810199253

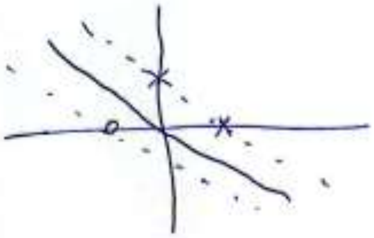
Spring 2022

فهرست

سوال ۱-۲-۳-۴

تمرین ۴ ML سامان کشاورز ۸۱۰۱۹۹۲۵۳

(1)



$\times$: class $(+1)$: $\begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}1\\0\end{pmatrix}$

$\circ$: class (-1) : $\begin{pmatrix}-1\\0\end{pmatrix}$

باید یک بعد بایاس اضافه کنیم تا معادلات به صورت مطلوب دربیاید.

$$s_1 = \begin{pmatrix}-1\\0\\1\end{pmatrix} \quad s_2 = \begin{pmatrix}0\\1\\1\end{pmatrix} \quad s_3 = \begin{pmatrix}1\\0\\1\end{pmatrix}$$

$$\begin{cases} \alpha_1 s_1^T s_1 + \alpha_2 s_2^T s_1 + \alpha_3 s_3^T s_1 = -1 & s_1 \text{ دسته اول} \\ \alpha_1 s_1^T s_2 + \alpha_2 s_2^T s_2 + \alpha_3 s_3^T s_2 = +1 & s_2 \text{ دسته دوم} \\ \alpha_1 s_1^T s_3 + \alpha_2 s_2^T s_3 + \alpha_3 s_3^T s_3 = +1 & s_3 \text{ دسته دوم} \end{cases}$$

$$\left.\begin{cases} 2\alpha_1 + \alpha_2 = -1 \\ \alpha_1 + 2\alpha_2 + \alpha_3 = +1 \\ \alpha_2 + 2\alpha_3 = +1 \end{cases}\right\} \Rightarrow \alpha_1 = -1 \quad \alpha_2 = 1 \quad \alpha_3 = 0$$

چون $\alpha_3 = 0$ شد پس s_3 جز SV نیست و فقط s_1 و s_2 SV هستند و برای رسم خطوط حاشیه کفایت می‌کنند.

$$\tilde{w} = \sum_{i=1}^{3} \alpha_i s_i = -1\begin{pmatrix}-1\\0\\1\end{pmatrix} + 1\begin{pmatrix}0\\1\\1\end{pmatrix} + 0\begin{pmatrix}1\\0\\1\end{pmatrix} = \begin{pmatrix}1\\1\\0\end{pmatrix} = \tilde{w}$$

$$\rightarrow y = wx + b \rightarrow w = \begin{pmatrix}1\\1\end{pmatrix} \quad \text{و } b = 0$$

② بخش اول:

$$K(x_i, x_j) = \exp\left(\frac{-1}{\alpha}\|x_i - x_j\|^2\right)$$

$$\|\varphi(x_i) - \varphi(x_j)\|^2 \leqslant k \longrightarrow$$ ثابت شود.

$$K(x_i, x_j) = \varphi^T(x_i)\,\varphi(x_j)$$

$$\|\varphi(x_i) - \varphi(x_j)\|^2 = (\varphi(x_i) - \varphi(x_j))^T(\varphi(x_i) - \varphi(x_j))$$

$$= \underbrace{\varphi^T(x_i)\varphi(x_i)}_{K(x_i,x_i)} - \underbrace{\varphi^T(x_i)\varphi(x_j)}_{K(x_i,x_j)} - \underbrace{\varphi^T(x_j)\varphi(x_i)}_{K(x_j,x_i)} + \underbrace{\varphi^T(x_j)\varphi(x_j)}_{K(x_j,x_j)}$$

$$\longrightarrow \underbrace{2\exp(0)}_{2} - 2\exp\left(\frac{-1}{\alpha}\|x_i - x_j\|^2\right) = \|\varphi(x_i) - \varphi(x_j)\|^2$$

$$\|\varphi(x_i) - \varphi(x_j)\|^2 = 2 - \underbrace{2\exp\left(\frac{-1}{\alpha}\|x_i - x_j\|^2\right)}_{max = 2}$$

$$\longrightarrow \boxed{\|\varphi(x_i) - \varphi(x_j)\|^2 \leqslant 2} \quad \boxed{K = 2}$$

بخش دوم:

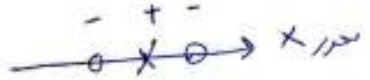
$$K(x_i, x_j) = \tanh(a\,x_i^T x_j + b) = \varphi^T(x_i)\varphi(x_j)$$

$$\|\varphi(x_i) - \varphi(x_j)\|^2 = (\varphi(x_i) - \varphi(x_j))^T(\varphi(x_i) - \varphi(x_j))$$

$$= \underbrace{\varphi^T(x_i)\varphi(x_i)}_{K(x_i,x_i)} - \underbrace{\varphi^T(x_i)\varphi(x_j)}_{K(x_i,x_j)} - \underbrace{\varphi^T(x_j)\varphi(x_i)}_{K(x_j,x_i)} + \underbrace{\varphi^T(x_j)\varphi(x_j)}_{K(x_j,x_j)}$$

$$\tanh(a\,x_i^T x_i + b) - \tanh(a\,x_i^T x_j + b) - \tanh(a\,x_j^T x_i + b) + \tanh(a\,x_j^T x_j + b)$$

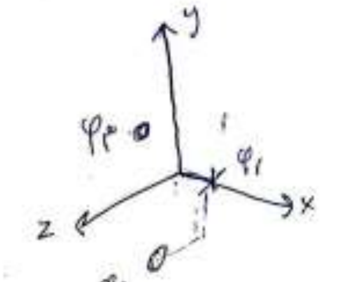
ماکسیمم مقدار $\tanh = 1$ است در نتیجه: $\|\varphi(x_i) - \varphi(x_j)\|^2 \leqslant 4$



③

الف) چون در یک بعد دو کلاس اصلا جدای پذیر نیستند.

ب) $\varphi(x_3) = \begin{bmatrix} 1 \\ \sqrt{2} \\ 1 \end{bmatrix}$ ، $\varphi(x_2) = \begin{bmatrix} 1 \\ -\sqrt{2} \\ 1 \end{bmatrix}$ ، $\varphi(x_1) = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$



چون مولفه x همه نقاط یکسان است، صفحه جدا کننده دارای بردار وزن $[0, 0, 0/5)$ باشد.

ج)

$$w = \sum_{i=1}^{3} \lambda_i y_i x_i$$

$$\min_{w,b} \frac{1}{2} \|w\|^2$$

$$\text{s.t.} \quad y_i (w^T \varphi(x) + b) \geq 1 , \quad i = 1, 2, 3$$

چون هر ۳ نقطه که داریم، support vetor هستند می‌توان محدودیت را با شرط مساوی حل کرد. تابع لاگرانژ را تشکیل دهیم:

$$L(w, b, \lambda) = \frac{1}{2} w^T w + \sum_{i=1}^{3} \lambda_i \left(1 - y_i \left(w^T \varphi(x_i) + b\right)\right)$$

حال باید نسبت به متغیرهای primal مشتق بگیریم:

$$\nabla_w L(w, b, \lambda) = w + \sum_{i=1}^{3} \lambda_i y_i \varphi(x_i) = 0$$

$$\nabla_b L(w, b, \lambda) = \sum_{i=1}^{3} \lambda_i y_i = 0$$

با جایگذاری $\varphi(x)$ ها خواهیم داشت:

$$\begin{cases} w_1 + \lambda_1 - \lambda_2 - \lambda_3 = 0 \\ w_2 + \sqrt{2}\lambda_2 - \sqrt{2}\lambda_3 = 0 \\ w_3 - \lambda_2 - \lambda_3 = 0 \\ \lambda_1 - \lambda_2 - \lambda_3 = 0 \end{cases} \longrightarrow w_1 = 0$$

با جایگذاری در معادلات:

$$b = 1 \quad -\sqrt{2}w_2 + w_3 + b = -1 \quad \sqrt{2}w_2 + w_3 + b = -1$$

$$\longrightarrow w_2 = 0 \text{ و } w_3 = -2$$

پس خواهیم داشت:

$$w = \begin{bmatrix} 0 \\ 0 \\ -2 \end{bmatrix} \quad b = 1$$

اندازه حاشیه ها با توجه به w و b، $0/5$ به دست می آید.

4) الف)

$$x_i . w + b \geqslant +1 - \varepsilon_i \quad y_i = +1 \quad \text{و} \quad \varepsilon_i \geqslant 0$$

$$x_i w + b \leqslant -1 + \varepsilon_i \quad y_i = -1$$

اگر x_i متعلق به $y_i = +1$ باشد و درست دسته بندی شده باشد: $0 \leq \varepsilon_i \leq 1$

اما اشتباه طبقه بندی انجام شده باشد:

$$y_i = -1 \quad \left.\begin{matrix} y_i(w^T x_i + b) < 0 \\ 1 - \varepsilon_i < 0 \end{matrix}\right\} \longrightarrow \varepsilon_i > 1$$

درحالتی که x_i متعلق به $y_i = -1$ باشد باز هم $\varepsilon \geqslant 1$ به دست می آید

$$\sum_{y_i \,(\text{misclassification})} \varepsilon_i > \sum 1 (\text{miss classification})$$

ادامه سوال 4 :

$$\sum_{\substack{i=\text{correct}\\ \text{classification}}} \varepsilon_i \geq 0$$

در نتیجه مجموع همه داده های طبقه بندی شده اشتباه برابر با حدبالای خطا خواهد شد.

ب)

$$K(x,y)^2 \leq K(x,x)K(y,y)$$

$$K(y,x) = \overline{K(x,y)}$$

باید ماتریس هسین مثبت معین باشد.

$$\begin{bmatrix} K(x,x) & \overline{K(x,y)} \\ K(x,y) & K(y,y) \end{bmatrix}$$

در نتیجه باید :

$$K(x,x)K(y,y) - K(x,y)^2 \geq 0$$

$$\rightarrow K^2(x,y) \leq K(x,x)K(y,y)$$

سوال ۵

الف)

کرنل‌ای svm توابعی هستند که ورودی با بعد کم را به فضایی معمولا با بعد زیاد منتقل می کنند تا بتوانند در فضای جدید الگوریتم دسته بندی را به خوبی اجرا کنند. توابع کرنل بیشتر در مسائل جداسازی غیرخطی مفید هستند که نمیتوانیم با یک خط ساده در فضای ورودی دسته بندی را انجام دهیم. با استفاده از این کرنلها ورودی به فضای جدید میرود که در ان داده ها بر اساس کلاسشان قابل جداسازی هستند. از جمله این توابع می توان به linear ، rbf، polyاشاره کرد و در حالت پیشفرض کرنل روی rbf قرار دارد. کرنل های rbf وpoly برای مسائل غیرخطی کاربرد بیشتری دارند. کرنل linear زمانی استفاده می شود که داده ها به صورت Linearly separable باشند، یعنی می توان آنها را با استفاده از یک خط جدا کرد. بیشتر زمانی استفاده می شود که تعداد زیادی ویژگی در یک مجموعه داده خاص وجود داشته باشد. کرنل RBF به دلیل شباهت آن به الگوریتم K-Nearest Neighborhood محبوب است. مزایای K-NN را دارد و بر مشکل پیچیدگی ابعاد زیاد غلبه میکند. زیرا فقط باید نقاط support vector را ذخیره کند نه کل داده را.

ب)

ابتدا توابع مورد نظر را فراخوانی میکنیم و سپس داده را بارگذاری میکنیم. در ادامه نیاز است که تبدییل شود که از دستور StandardScaler().fit_transform(X) استفاده میکنیم.

در ادامه هر ۴ کرنل داده شده را با C های خواسته شده در سوال پیاده سازی میکنیم. برای انجام این کار از حلقه for استفاده کردم. برای نمونه یکی از کدهای درون یک حلقه مربوط به کرنل rbf به صورت زیر است:

```
for i in C_range:
    model = SVC(kernel="rbf" , C=i)
    model.fit(X_train, y_train)
    predictions_train = model.predict(X_train)
    predictions_test = model.predict(X_test)
    accuracy_train = accuracy_score(y_train, predictions_train)
    accuracy_test = accuracy_score(y_test, predictions_test)
    print(f'train C{i} : {accuracy_train}')
    print(f'test C{i} : {accuracy_test}')
```

این کد از روش svm با کرنل مذکور ساتفاده میکنید و دقت داده تست و اموزش را در نهایت نمایش خواهد داد. نمونه اجرای همین تکه کد به صورت زیر است:

Rbf:

```
train C1 : 0.979007024265645
test C1 : 0.9798882681564246
train C100 : 0.9829980842911877
test C100 : 0.9824953445065177
train C1000 : 0.9857918263090677
test C1000 : 0.9810055865921787
```

برای بقیه کرنلها هم مقادیر به صورت زیر خواهد بود:

Linear:

```
train C1 : 0.978448275862069
test C1 : 0.9819366852886406
train C100 : 0.9786877394636015
test C100 : 0.9815642458100559
train C1000 : 0.9786079182630907
test C1000 : 0.9815642458100559
```

sigmoid:

```
train C1 : 0.8706896551724138
test C1 : 0.873929236499069
train C100 : 0.8706896551724138
test C100 : 0.8735567970204842
```

poly:

```
train C1 : 0.9761334610472542
test C1 : 0.9795158286778398
train C100 : 0.9798052362707536
test C100 : 0.9806331471135941
```

ت)

بهترین طبقه بند قسمت ب rbf با c=100 است که البته کرنل خطی با c=1 بسیار عملکرد نزدیکی به آن دارد. نتیجه اعمال ان بر داده و دقت و ماتریس کانفیوژن در ادامه امده است:

```
test  : 0.9824953445065177
              precision    recall  f1-score   support

           0       0.99      0.99      0.99      4899
           1       0.94      0.86      0.90       471

    accuracy                           0.98      5370
   macro avg       0.96      0.93      0.94      5370
weighted avg       0.98      0.98      0.98      5370
```

سوال ۶

الف)

وقتی که چندین مدلهای یادگیر ضعیف برای حل یک مسئله آموزش داده می شوند و برای ساخت یک مدل یادگیرنده قوی تر ترکیب شوند، مدل های یادگیری ماشین آنسمبل ساخته میشود. ترکیب صحیح و اصولی مدل های ضعیف مدل های پایدارتر و بهتری میسازند. داشتن مدلی با بایاس و واریانس کم دو ویژگی اساسی و مطلوب می باشند که شاید به کمک انسمبل کردن مدلهای ضعیف بتوان به ان دست یافت. در مدل های یادگیری ماشین، انتخاب الگوریتم ها در به دست آوردن نتایج خوب بسیار مهم است. متغیرهای زیادی در مسئله مانند مقدار داده ها، ابعاد داده ها و فرضیه توزیع درانتخاب معدل دخیل هستند. بیشتر اوقات این مدل های ضعیف به تنهایی عملکرد خوبی ندارند زیرا دارای بایاس یا به خصوص واریانس زیاد هستند .

به کمک انسمبل کردن مدلها از چندین مدل یادیری مختلف استفاده میکنیم که ممکن است هر کدام مزایا خود را برای تصمیم گیری داشته باشند. سپس بین نتایج مدلهای مختلف میتوانیم رای گیری کنیم. معمولا این رای گیری به صورت وزندار انجام میشود و وزن هر کلاس انتخاب شده توسط مدل متناسب با دقت آن مدل تصمیم گیری است. گاهی هم میانگین وزندار میان نتایج گرفته میشود که باز هم از نوع رای گیری است.