

به نام خدا  
دانشگاه صنعتی امیرکبیر  
(پلی تکنیک تهران)  
دانشکده مهندسی کامپیوتر



دانشگاه صنعتی امیرکبیر  
(پلی تکنیک تهران)

## گزارش تمرین سوم درس یادگیری ماشین

استاد درس: دکتر احسان ناظرفرد

دانشجو: فاطمه غلامزاده

۹۹۱۳۱۰۰۳

نیم سال دوم ۱۳۹۹-۱۴۰۰

## سوالات تشریحی

### سوال ۱:

الف) بله. چون در این الگوریتم سعی می‌شود مقادیر متناظر با خط جداکننده پیدا شود. این خط نیز به توزیع داده‌ها بستگی دارد. بنابراین در الگوریتم SVM یادگیری بر توزیع داده‌هاست و پارامترهای خط به توزیع داده‌ها بستگی دارد. این توضیحات مربوط به SVM پایه است. در مورد SVM های غیرخطی این مطلب صحیح نیست و غیرپارامتری هستند.

ب) بله. حاشیه‌ی هر مدل میزان عملکرد آن مدل در مقابل داده های تست را نشان می‌دهد. هرچه این حاشیه بیشتر باشد مدل بهتر است و کمتر دچار اشتباه در تصمیم‌گیری می‌شود.

ج) خیر. در SVM پایه که hard است بردارهای پشتیبان شدیداً به داده‌ها حساس است و دچار بیش برازش می‌شود ولی در نسخه‌ی soft آن می‌توان با ضریب بتا این میزان بیش برازش شدن را کاهش داد.

د) خیر. داده‌های نویزی و پرت در تعیین حاشیه در حالت سخت (SVM hard) تاثیر دارد. ولی در حالت نرم (Soft SVM) میتوان تاثیر این داده ها را با کم در نظر گرفتن میزان جریمه، کاهش داد.

### سوال ۲:

فرمول محاسبه وزن هر دسته‌بند:

$$\alpha^t = \frac{1}{2} \ln \frac{1 - \epsilon^t}{\epsilon^t}$$

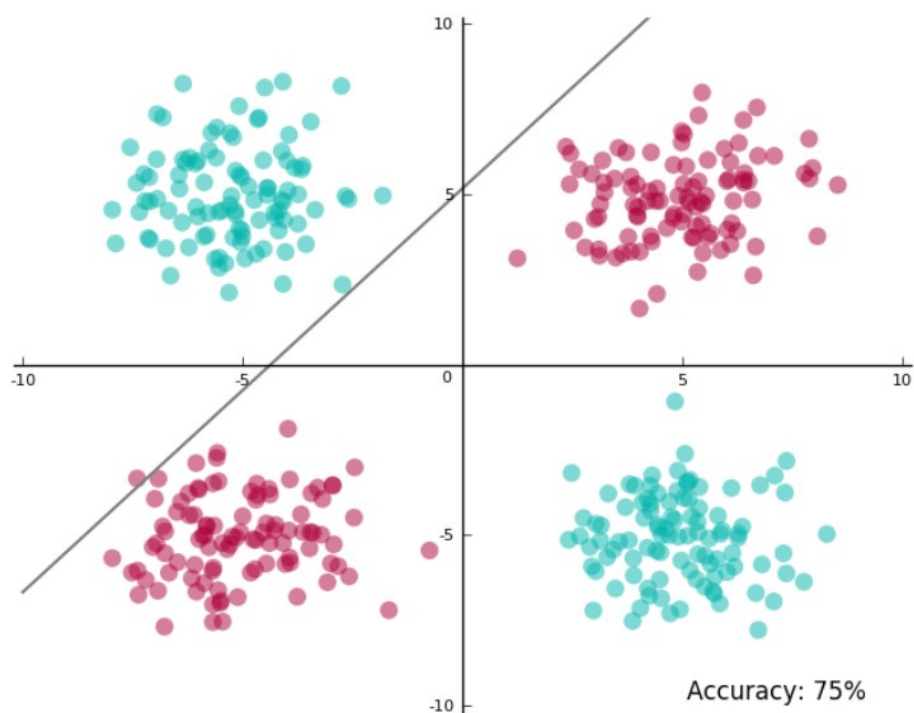
چون در نتیجه‌ی اجرای h1 فقط یک داده به اشتباه دسته بندی شده است خطا برابر یک هشتم است:

$$e_1 = 1/8$$

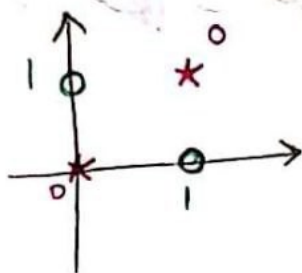
$$\alpha_1 = \frac{1}{2} * \ln \left( \frac{1-1/8}{1/8} \right) = \frac{1}{2} * \ln(7) = 0.97$$

### سوال ۳:

بله در این مسئله امکان استفاده از SVM را داریم. این مسئله در واقع مسئله‌ی دسته‌بندی XOR است و با SVM قابل حل است. اگر از کرنل خطی استفاده کنیم در بهترین حالت دقت ۷۵ درصد داریم که به شکل زیر است:



برای دقت بالاتر، باید از کرنل چندجمله‌ای درجه ۲ استفاده شود. به عنوان مثال مسئله XOR ساده که دارای ۴ نقطه است را با کرنل درجه ۲ حل می‌کنیم:



train set:

دیتای آموزش:

$x_1$	$x_2$	$y$
0	0	0
1	1	0
1	0	1
0	1	1

$$\phi(x) = (x_1^2, \sqrt{2}x_1x_2, x_2^2)$$

از کرنل درجه ۲ استفاده می کنیم:

$x_1$	$x_2$	$y$	$\phi_1$	$\phi_2$	$\phi_3$	
0	0	0	0	0	0	$\rightarrow x'_1$
1	1	0	1	$\sqrt{2}$	1	$\rightarrow x'_2$
1	0	1	1	0	0	$\rightarrow x'_3$
0	1	1	0	0	1	$\rightarrow x'_4$

بنابراین دیتا سِت جدید ما (در فضای سه بعدی) شامل ۴ نقطه  $x'_1$ ،  $x'_2$ ،  $x'_3$  و  $x'_4$  است.

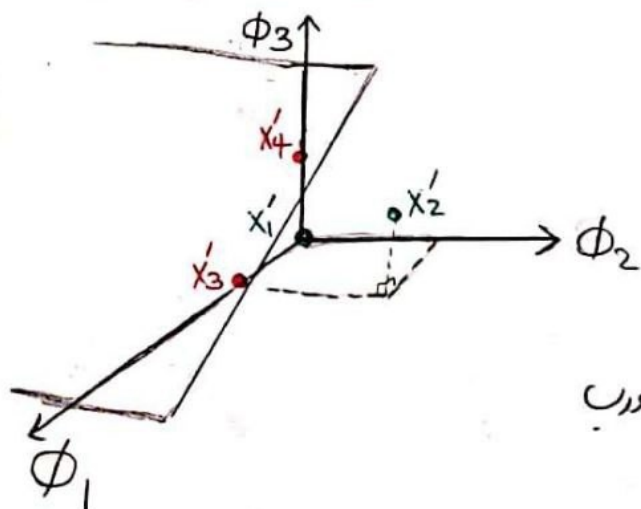
این ۴ نقطه را در فضای ۳ بعدی می بینیم:

با توجه به شکل به دست آمده این ۴ نقطه در فضای ۳ بعدی به راحتی توسط یک صفحه

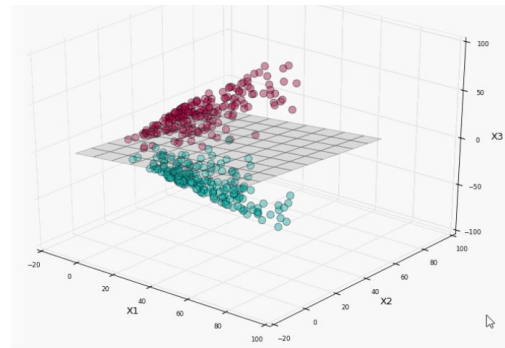
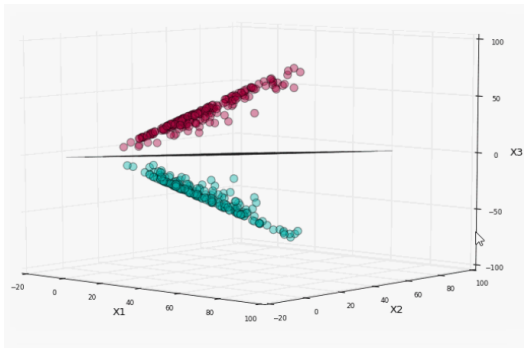
عباز پذیرند. مثلاً صفحه ای که در شکل نشان

داره شده و صفحه  $\phi_1$ ،  $\phi_3$  را در یک خط مموم

قطع کرده است.



در واقع در فضای سه بعدی نقاط به این شکل در می آیند و با صفحه از هم جدا می شوند:



#### سوال ۴:

الف) بله الگوریتم آدابوست نسبت به داده نویزی حساس است زیرا این الگوریتم روی خود داده ها کار می کند و به خود داده ها اهمیت زیادی می دهد و به آن ها وزن اختصاص می دهد. چنین الگوریتم هایی را **instance based** می گویند و این الگوریتم ها به نویز خیلی حساس هستند. از طرفی تابع هزینه به صورت نمایی رشد می کند و اگر داده نویز در داده ها باشد پنالتی زیادی را برای این داده در نظر می گیرد اما با این وجود این مورد باعث بیش برآزش این الگوریتم نمی شود و بر روی دقت الگوریتم روی داده های تست تاثیر چندانی ندارد.

ب) خیر. اگر این دسته بند ضعیف یک دسته بند تصادفی باشد در هر اجرا از داده های آموزش برچسب مختلف به داده ها می زند و اصلا رفتار قابل پیش بینی ای ندارد. ولی اگر یک دسته بند معقول و غیر تصادفی باشد این گزاره صحیح است.

## سوالات پیاده سازی

مسئله ۱) پیاده سازی ماشین بردار پشتیبان

ب) درصد داده های تست و ترین : 25% , 75%

کرل خطی:

	C=1	C=0.01	C=0.001
Accuracy	0.81	0.90	0.75
F1-measure	0.72	0.84	0.2

کرل چند جمله ای:

	C=1000 , d= 4	C=100 , d= 3	C=100,d=5
Accuracy	0.62	0.84	0.78
F1-measure	0.30	0.70	0.36

کرل RBF:

	Gamma = auto	Gamma=0.001	Gamma=0.002
Accuracy	0.93	0.87	0.90
F1-measure	0.87	0.71	0.8

کرل sigmoid :

	C=1	C=1.5	C=2
Accuracy	0.84	0.90	0.87
F1-measure	0.73	0.85	0.81

(ج)

- کرنل خطی تنها یک پارامتر  $C$  دارد که با تغییر دادن آن تنها بایاس مدل تغییر می‌کند و تاثیری در شکل مدل ندارد.
- کرنل چندجمله‌ای دو پارامتر درجه و  $C$  دارد که با افزایش پارامتر  $d$ ؛ یعنی درجه، میزان پیچیدگی مدل افزایش میابد و با زیاد شدن هرچه بیشتر آن بیش برازش بیشتری صورت می‌گیرد. تاثیر  $C$  نیز مانند حالت قبل است.
- کرنل RBF یک پارامتر گاما دارد که هرچه گاما بیشتر باشد پیچیدگی مدل بیشتر است و در نتیجه مدل بر داده‌ها منطبق می‌شود و بیش برازش رخ می‌دهد.
- کرنل سیگنویید یک پارامتر  $C$  دارد که این پارامتر نیز با توجه به آزمایش‌ها بیشتر بودنش بهتر است. ولی از حدی نباید بیشتر شود. چون عملکرد کاهش می‌یابد.

(د) بهترین مدل: کرنل RBF با مقدار  $\text{gamma} = \text{'auto'}$

## مسئله ۲) بررسی عملکرد مدل‌های ترکیبی

(الف)

accuracy : 0.73

ماتریس درهم‌ریختگی:

		Predicted Class									
True Class		0	1	2	3	4	5	6	7	8	9
	0	321	1	0	0	0	0	15	9	17	0
	1	0	143	214	5	2	0	0	0	0	0
	2	0	1	351	3	1	0	2	6	0	0
	3	0	8	0	326	1	0	0	1	0	0
	4	0	0	0	2	358	0	0	0	0	4
	5	0	2	0	111	18	166	5	0	2	31
	6	5	0	4	5	0	0	307	13	2	0

7	0	34	8	6	22	3	5	258	1	0
8	18	0	0	0	0	26	3	44	244	0
9	0	59	0	90	50	0	2	0	1	134

(ب)

accuracy : 0.9502430654847012

(ج)

	#Decision Tree = 5	#Decision Tree= 20	#Decision Tree= 50
accuracy	0.93937	0.96225	0.963969

(د)

پارامترهای بهینه:

```
learning_rate =0.1,
n_estimators=1000,
max_depth=5,
min_child_weight=1,
gamma=0,
subsample=0.8,
colsample_bytree=0.8,
reg_lambda=0.1,
objective= 'logistic',
nthread=-1,
scale_pos_weight=1,
booster='gbtree',
seed=77
```

accuracy : ۰٫۹۶۳۹۶