دانشکده‌ی مهندسی کامپیوتر

یادگیری ماشین

نیم‌سال دوم ۱۴۰۲-۱۴۰۱

مدرس: دکتر سید ابوالفضل مطهری

تمرین چهارم

مسئله‌ی ۱. Boosting

(آ) الگوریتم AdaBoost به دنبال کمینه کردن تابع خطای نمایی زیر است:

$$E = \sum_{i=۱}^{N} \exp(-y_i f(x_i))$$

که در آن $y = +۱$ یا $y = -۱$ برچسب دسته، x داده، و $f(x)$ جمع وزن‌دار weak learner ها هستند. نشان دهید که E که به صورت اکید بزرگ‌تر از تابع L بوده و در نتیجه یک حد بالا برای L است که بصورت زیر تعریف می‌شود:

$$L = \sum_{i=۱}^{N} ۱ \times (y_i f(x_i) < ۰).$$

(ب) در الگوریتم AdaBoost در دو مورد باید احتیاط کرد. در مورد آن‌ها، به دو مورد زیر پاسخ دهید.

- به صورت ریاضی، نشان دهید که چرا weak learner با دقت[۱] کمتر از ۵۰ درصد، برای AdaBoost مشکل ایجاد می‌کند.
- الگوریتم AdaBoost به داده‌های پرت حساس است. یک پیشنهاد ساده ابتکاری[۲] ارائه دهید که این مشکل را تا حدی برطرف کند.

مسئله‌ی ۲.

می‌خواهیم یک درخت تصمیم[۳] را روی دیتاست زیر آموزش دهیم تا بتوانیم دانش‌آموزان تنبل و زرنگ را از هم تشخیص دهیم. برای اینکار از سه ویژگی استفاده می‌کنیم:

---

[۱] accuracy
[۲] heuristic
[۳] decision tree

G1	G2	G3	Output
N	A	2	L
N	V	2	L
N	V	2	L
U	V	3	L
U	V	3	L
U	A	4	D
N	A	4	D
N	V	4	D
U	A	3	D
U	A	3	D

۱. آنتروپی شرطی زیر را حساب کنید.

$H(G2|G1 = N)$

۲. الگوریتم ID3 چه ویژگی‌ای را به عنوان ریشه درخت انخاب می‌کند؟

۳. درخت تصمیم آموزش دیده روی این داده‌ها را رسم کنید.

مسئله‌ی ۳.

برای داده های جدول‌های ۱ و ۲ درخت تصمیم را به صورت دستی بر روی داده های یادگیری آموزش دهید و دقت دسته‌بند را بر روی داده‌های تست بررسی نمایید.

Target	G5	G4	G3	G2	G1	Name
+	High	Low	High	Low	Low	X1
–	Low	Low	Medium	Low	High	X2
+	High	Low	High	High	Low	X3
–	Low	Low	Low	Low	High	X4
+	Low	High	Medium	High	Low	X5
–	High	High	Medium	Low	Low	X6
–	High	High	High	High	Low	X7
+	Low	Low	Low	Low	Low	X8
–	Low	High	Medium	High	High	X9

جدول ۱: داده های یادگیری

Target	G5	G4	G3	G2	G1	Name
+	High	Low	High	Low	Low	X1
–	High	High	Low	High	High	X2
+	High	Low	High	Low	Low	X3
–	High	Low	Medium	Low	Low	X4
+	High	Low	Medium	High	Low	X5
–	Low	Low	High	High	Low	X6

جدول ۲: داده های تست

مسئله‌ی ۴.

فرض کنید داده‌های یک بعدی $D = \{۴, ۱, ۹, ۱۲, ۶, ۱۰, ۲, ۳, ۹\}$ داده شده‌اند. می‌خواهیم الگوریتم k-means را برای $k = ۲$ اجرا کنیم.

الف) فرض کنید در ابتدا به صورت تصادفی مرکز خوشه‌ها به ترتیب ۱ و ۶ هستند. بنابراین در ابتدا داریم:

$$\text{Cluster } ۱ = \{۱, ۲, ۳\}\,,\ \text{Cluster } ۲ = \{۴, ۹, ۱۲, ۶, ۱۰, ۹\}$$

الگوریتم را برای ۲ مرحله دیگر ادامه دهید. در هر مرحله، خوشه‌ها و مراکز خوشه‌ها را مشخص کنید.

ب) بعد از این دو مرحله آیا الگوریتم همگرا شده است؟ دلیل خود را بنویسید.

مسئله‌ی ۵.

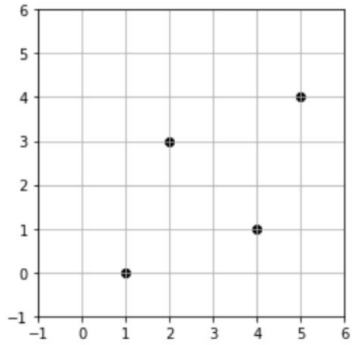
ماتریس داده زیر را در نظر بگیرید که شامل ۴ داده ۲ ـ بعدی است.

$$X = \begin{bmatrix} ۴ & ۱ \\ ۲ & ۳ \\ ۵ & ۴ \\ ۱ & ۰ \end{bmatrix}$$

می‌خواهیم به کمک PCA ابعاد داده‌ها را کاهش بدهیم؛ به طوری که یک‌ـ بعدی شوند.

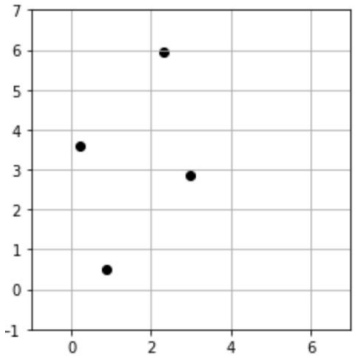
الف) بردارهای واحد principal component را برای X حساب کنید. الگوریتم PCA کدام یک را انتخاب می‌کند؟ (جواب را دقیق به دست آورید و از روش‌های تخمینی استفاده نکنید).

ب) ۴ نقطه داده شده را در فضای دوـ بعدی رسم کنید. (شکل ۱) سپس principal component را ترسیم کرده و نقاط را بر روی آن تصویر کنید. برای هر نقطه تصویر شده principal coordinate value را مشخص کنید.



شکل ۱: نقاط X در صفحه

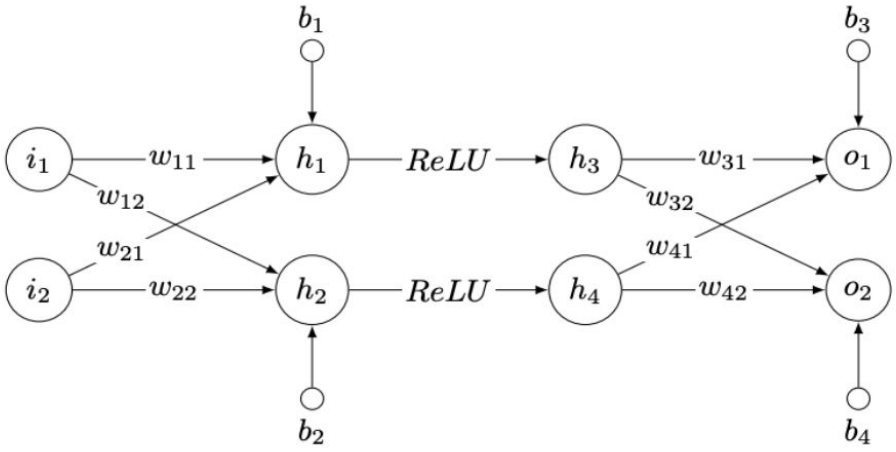
پ) شکل ۲ نقاط X را نشان می‌دهد که ۳۰ درجه خلاف جهت عقربه‌های ساعت چرخیده‌اند. مشابه دو قسمت قبلی، برای این نقاط هم PCA را اجرا کرده و تصویر نقاط را روی principal direction رسم کنید. همچنین دوباره برای هر کدام principal coordinate value را بنویسید. از جواب چه نتیجه‌ای می‌گیرید؟



شکل ۲: نقاط X بعد از ۳۰ درجه چرخش خلاف جهت عقربه‌های ساعت

مسئله‌ی ۶.

(امتیازی) شبکه عصبی شکل ۳ را با تابع فعال‌سازی ReLU در نظر بگیرید. (i_1, i_2) ورودی هستند، دو لایه مخفی داریم و خروجی‌ها در انتها (o_1, o_2) هستند. برچسب داده‌ها با (t_1, t_2) ، وزن‌ها با w و بایاس با b نشان داده شده است.



شکل ۳: شبکه عصبی

مقادیر متغیرها را هم می‌توانید در جدول شکل ۴ مشاهده کنید.

Variable	i_1	i_2	w_{11}	w_{12}	w_{21}	w_{22}	w_{31}	w_{32}	w_{41}	w_{42}	b_1	b_2	b_3	b_4	t_1	t_2
Value	2.0	-1.0	1.0	-0.5	0.5	-1.0	0.5	-1.0	-0.5	1.0	0.5	-0.5	-1.0	0.5	1.0	0.5

شکل ۴: جدول مقادیر متغیرها

الف) خروجی (o_1, o_2) را با توجه به مقادیر داده شده به دست آورید. تمامی محاسبات را بنویسید.

ب) خطای MSE را حساب کنید.

پ) فرض کنید تابع هزینه همان قسمت (ب) باشد. مقدار وزن $w_{۲۱}$ را با کمک gradient descent با نرخ یادگیری ۰/۱ آپدیت کنید. (تمامی محاسبات را بنویسید).