به نام خدا

دانشکده مهندسی کامپیوتر

درس یادگیری ماشین
نیم‌سال دوم ۰۳-۰۴
استاد: ربیعی- دهقانیان

آزمونک چهارم - ۱۴۰۴/۰۲/۲۱ زمان آزمون: ۲۰ دقیقه

۱. (۵۰ نمره) فرض کنید شش نقطه‌ی آموزشی (در جدول ۱) برای یک مسئله‌ی دسته‌بندی در اختیار دارید که شامل دو ویژگی دودویی X_1 و X_2، و سه کلاس $Y \in \{1, 2, 3\}$ است. ما از قرار است از درخت تصمیم استفاده کنیم.

X_1	X_2	Y
1	1	1
1	1	1
1	1	2
1	0	3
0	0	2
0	0	3

جدول ۱: داده‌های آموزشی

(آ) مقدار کسب اطلاعات (Information-Gain) را برای هر دو ویژگی X_1 و X_2 محاسبه کنید. می‌توانید از تقریب $\log_2 3 \approx \frac{19}{12}$ استفاده کنید.

حل

فرمول کلی بهره اطلاعات:

$$IG(X) = H(Y) - H(Y \mid X)$$
$$H(X) = -\sum_x P(X=x) \log_2 P(X=x)$$
$$H(Y \mid X) = \sum_x P(X=x) \sum_y P(Y=y \mid X=x) \log_2 P(Y=y \mid X=x)$$

برای محاسبه بهره اطلاعات هر تقسیم ابتدا آنتروپی $H(Y)$ را محاسبه می‌کنیم:

$$H(Y) = -\sum_{y_i=1}^{n=3} P(Y=y_i) \log_2 P(Y=y_i)$$
$$= -\sum_{y_i=1}^{n=3} \frac{1}{3} \log_2 \frac{1}{3} = \log_2 3 \approx \frac{19}{12}$$

و برای تقسیم‌بندی X_1، آنتروپی شرطی آن به صورت زیر محاسبه می‌شود:

$$
\begin{aligned}
H(Y \mid X_1) = -P(X_1 = 0) \sum_{y_i=1}^{n=3} P(Y = y_i \mid X_1 = 0) \log_2 P(Y = y_i \mid X_1 = 0) \\
-P(X_1 = 1) \sum_{y_i=1}^{n=3} P(Y = y_i \mid X_1 = 1) \log_2 P(Y = y_i \mid X_1 = 1) \\
= -\Big[\frac{2}{6}\left(\frac{0}{2}\log_2\frac{0}{2} + \frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right) \\
+ \frac{4}{6}\left(\frac{2}{4}\log_2\frac{2}{4} + \frac{1}{4}\log_2\frac{1}{4} + \frac{1}{4}\log_2\frac{1}{4}\right)\Big] \\
= -\left(-\frac{2}{6} - 1\right) = \frac{4}{3}
\end{aligned}
$$

به‌طور مشابه برای تقسیم‌بندی X_2 آنتروپی شرطی را محاسبه می‌کنیم:

$$
\begin{aligned}
H(Y \mid X_2) = -P(X_2 = 0) \sum_{y_i=1}^{n=3} P(Y = y_i \mid X_2 = 0) \log_2 P(Y = y_i \mid X_2 = 0) \\
-P(X_2 = 1) \sum_{y_i=1}^{n=3} P(Y = y_i \mid X_2 = 1) \log_2 P(Y = y_i \mid X_2 = 1) \\
= -\Big[\frac{3}{6}\left(\frac{0}{3}\log_2\frac{0}{3} + \frac{1}{3}\log_2\frac{1}{3} + \frac{2}{3}\log_2\frac{2}{3}\right) \\
+ \frac{3}{6}\left(\frac{2}{3}\log_2\frac{2}{3} + \frac{1}{3}\log_2\frac{1}{3} + \frac{0}{3}\log_2\frac{0}{3}\right)\Big] \\
\approx -\left(\frac{2}{3} - \frac{19}{12}\right) \\
= \frac{11}{12}
\end{aligned}
$$

در نتیجه، بهره اطلاعات (Information-Gain) برای هر تقسیم به صورت زیر خواهد بود:

$$
\begin{aligned}
IG(X_1) = H(Y) - H(Y \mid X_1) \approx \frac{19}{12} - \frac{4}{3} = \frac{19}{12} - \frac{16}{12} = \frac{1}{4} \\
IG(X_2) = H(Y) - H(Y \mid X_2) \approx \frac{19}{12} - \frac{11}{12} = \frac{8}{12} = \frac{2}{3}
\end{aligned}
$$

(ب) ابتدا بهتر است داده ها را براساس کدام ویژگی تفکیک کنیم؟

حل

چون مقدار information-gain ویژگی X_2 بیشتر است پس بهتر است از ان برای تفکیک استفاده کنیم.

۲. (۵۰ نمره) به یاد بیاورید که مسئله اولیه‌ی SVM با حاشیه‌ی نرم به صورت زیر است:

$$\min \quad \frac{1}{2}w^T w \;+\; C\sum_{i=1}^{n}\xi_i,$$
$$s.t. \quad \xi_i \geq 0, \quad \forall i \in \{1,\dots,n\},$$
$$(w^T x_i + b)\, y_i \geq 1 - \xi_i, \quad \forall i \in \{1,\dots,n\}.$$

برای به‌دست آوردن SVM با کرنل، دوگان این مسئله گرفته می‌شود و سپس $x_i^T x_j$ با $k(x_i, x_j)$ جایگزین می‌شود، که $k(\cdot,\cdot)$ تابع کرنل است.

شکل ۱ مرزهای تصمیم‌گیری SVM را با کرنل‌ها و/یا ضرایب جریمه‌ی متفاوت نشان می‌دهد. در این شکل، دو دسته داده‌ی آموزشی با برچسب‌های $y_i \in \{-1, 1\}$ داریم که با دایره و مربع مشخص شده‌اند. دایره‌ها و مربع‌های توپر، بردارهای پشتیبان (Support-Vectors) هستند. هر نمودار را با یکی از بهینه‌سازی‌های زیر مطابقت دهید (به طور خیلی مختصر توضیح دهید).

آ) یک SVM خطی با حاشیه‌ی نرم و $C = 0.1$

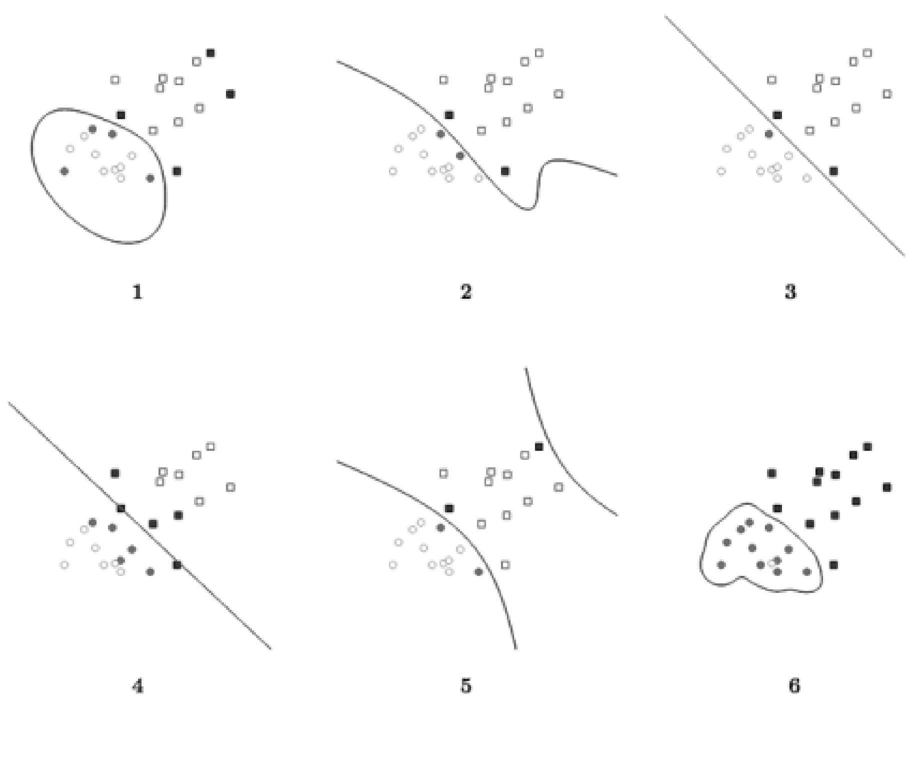
ب) یک SVM خطی با حاشیه‌ی نرم و $C = 10$

ج) یک SVM با حاشیه‌ی سخت و کرنل $K(u,v) = u^T v + (u^T v)^2$

د) یک SVM با حاشیه‌ی سخت و کرنل $K(u,v) = \exp\Big(-\frac{1}{4}\|u-v\|^2\Big)$

ه) یک SVM با حاشیه‌ی سخت و کرنل $K(u,v) = \exp(-4\|u-v\|^2)$

و) هیچ‌کدام از موارد بالا



شکل ۱:

حل

پاسخ‌ها:

(الف): ۴، (ب): ۳، (ج): ۵، (د): ۱، (ه): ۶، (و): ۲

دلیل:

(الف) مرز تصمیم SVM خطی است. در مقایسه با شکل ۳ (که مربوط به (ب) است)، خط به‌طور دقیق دو کلاس را جدا نمی‌کند، که این حالت مربوط به زمانی است که مقدار C کوچک باشد و خطاهای بیشتری مجاز باشند.

(ب) مرز تصمیم SVM خطی است. در مقایسه با شکل ۴ (که مربوط به (الف) است)، خط دو کلاس را به‌طور دقیق جدا می‌کند، که این حالت نشان‌دهنده مقدار بزرگ C است.

(ج) تابع تصمیم کرنل درجه دوم به صورت زیر تعریف می‌شود: $f(x) = \sum_i \alpha_i (x_i \cdot x + (x_i \cdot x)^2) + b$ در نتیجه، مرز تصمیم برابر است با $f(x) = 0$. از آنجا که $f(x)$ تابعی از مرتبه دوم نسبت به x است، منحنی می‌تواند بیضوی یا هایپربولیک باشد. شکل ۵ یک منحنی هایپربولیک را نشان می‌دهد.

(د) می‌توان تابع تصمیم را به صورت زیر نوشت:

$f(x) = \sum_i \alpha_i \exp(-\gamma \|x_i - x\|^2) + b$ اگر γ بزرگ باشد، مقدار کرنل حتی با وجود فاصله کم بین x و x_i بسیار کوچک می‌شود. این موضوع باعث می‌شود که طبقه‌بندی سخت و تعداد بردارهای پشتیبان کم باشد. اگر شکل ۱ مربوط به حالتی با $\gamma = 4$ (بزرگ) باشد، در آن صورت طبقه‌بندی نقاط دایره‌ای در وسط دشوار خواهد بود. بنابراین، شکل ۱ مربوط به $\gamma = 4$ است.

(ه) با استدلال مشابه، می‌توان نتیجه گرفت که اگر γ کوچک باشد (مثلاً $\gamma = \frac{1}{4}$)، بردارهای پشتیبان بیشتری وجود خواهند داشت.