



مقدمه‌ای بر یادگیری ماشین

پاییز ۱۴۰۱

اساتید: علی شریفی، بهروز آذرخلیلی

دانشگاه صنعتی شریف

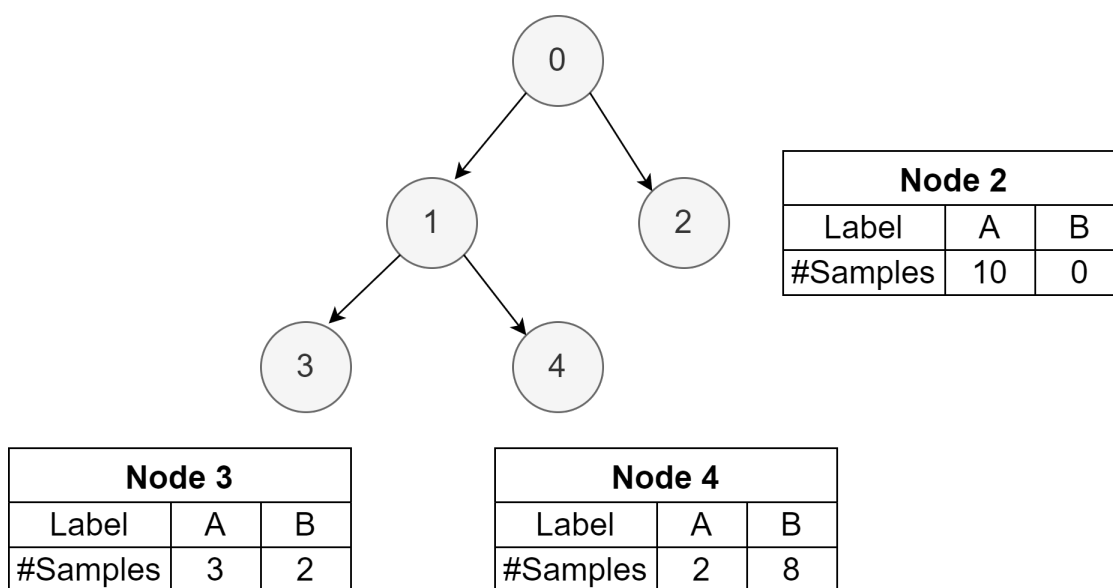
دانشکده‌ی مهندسی کامپیوتر

تاریخ برگزاری: ۲۰ آذر

میانترم

سوالات (۱۰۰ نمره)

۱. (۲۰ نمره) یک درخت تصمیم بر اساس نمونه‌های موجود برای دو کلاس A و B به صورت زیر ایجاد شده است که در آن تعداد نمونه‌ها در هر کلاس برای هر گره نشان داده شده است.



شکل ۱

با توجه به این درخت تصمیم به سوالات زیر پاسخ دهید:

الف) آنتروپی را برای این درخت تصمیم بر اساس گره ۰ محاسبه کنید.

ب) میزان اطلاع بدست آمده بر اساس گره ۰ را محاسبه کنید.

ج) معیار دقت را برای کلاس بندی صورت گرفته بر اساس این درخت تصمیم محاسبه کنید.

د) چگونه می‌توان دقت طبقه‌بندی را برای درخت تصمیم موجود افزایش داد؟ پیشنهاد خود را ارائه دهید.

ه) با راه‌حل ارائه شده برای قسمت قبل ممکن است با چه ریسکی در مورد این درخت تصمیم مواجه شویم و راهکار رهایی از این مشکل چیست؟

۲. (۲۰ نمره) مفاهیم زیر را بطور کامل توضیح دهید:

الف) در یک الگوریتم یادگیری ماشین تفاوت Validation Set و Test Set چیست؟

ب) چه تفاوتی در بکارگیری معیارهای محاسبه خطا MSE و MAE برای ارزیابی تکنیک‌های مختلف الگوریتم‌های یادگیری ماشین وجود دارد.

ج) تکنیک K-Fold Cross Validation را توضیح دهید. استفاده از آن را چه زمانی مناسب می‌دانید.

د) اگر بخواهیم اعداد ۱ تا ۱۵ را به عنوان خروجی یک نورون تولید کنیم، کدام تابع فعال‌ساز را مناسب می‌دانید؟ دلیل خود را بنویسید.

۳. (۲۰ نمره) شبکه داده شده در شکل زیر را در نظر بگیرید که تابع فعال‌ساز آن به صورت

$$f(z) = \frac{1}{1 + e^{-\|z\|^2}}$$

است، که در آن

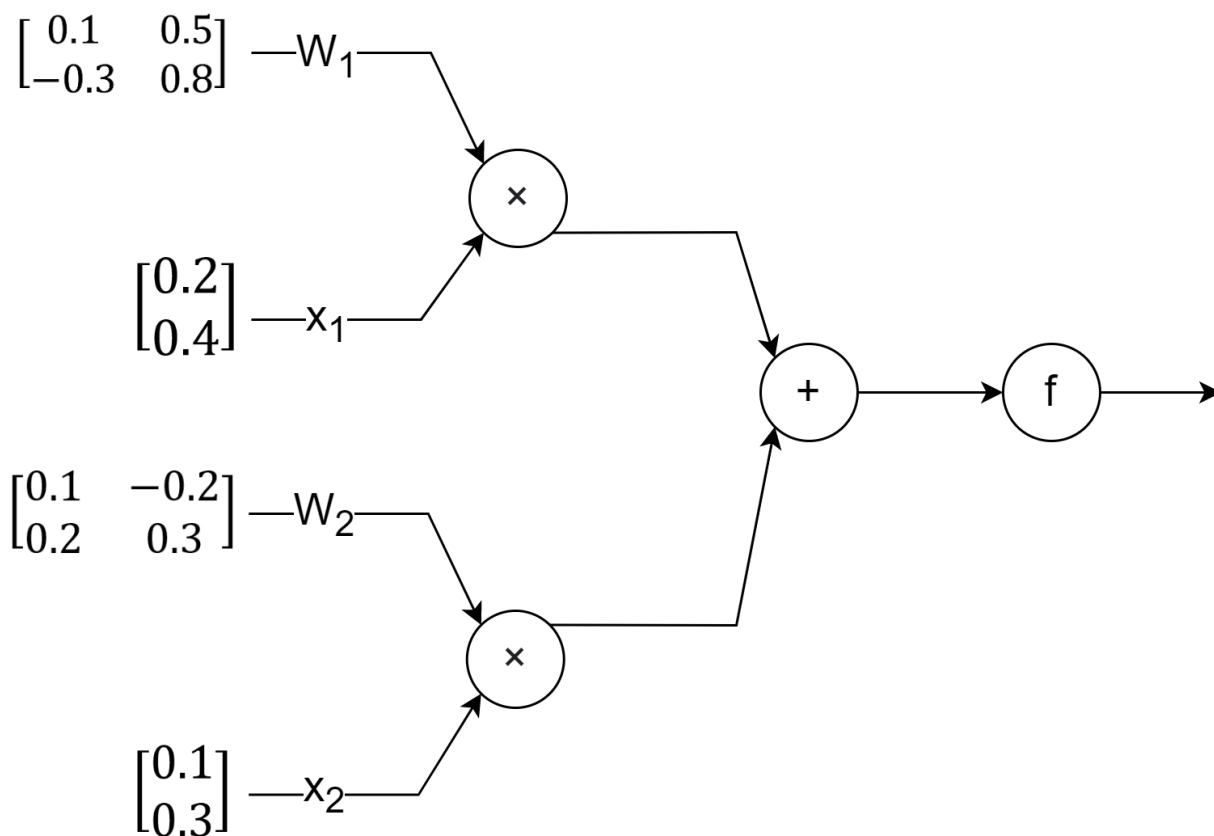
$$z = W_1 \cdot x_1 + W_2 \cdot x_2$$

است.

الف) مراحل محاسبه تابع فعال‌سازی (f) را به تفکیک به شکل مربوط به شبکه اضافه کنید.

ب) feed forward را برای شبکه بدست آمده از قسمت الف محاسبه کنید.

ج) مشتقات جزئی هر مرحله را با توجه به جواب‌های قسمت ب محاسبه کنید. (قابل توجه است که $\|z\|^2$ معادل توان ۲ نرم l_2 برای z است.)



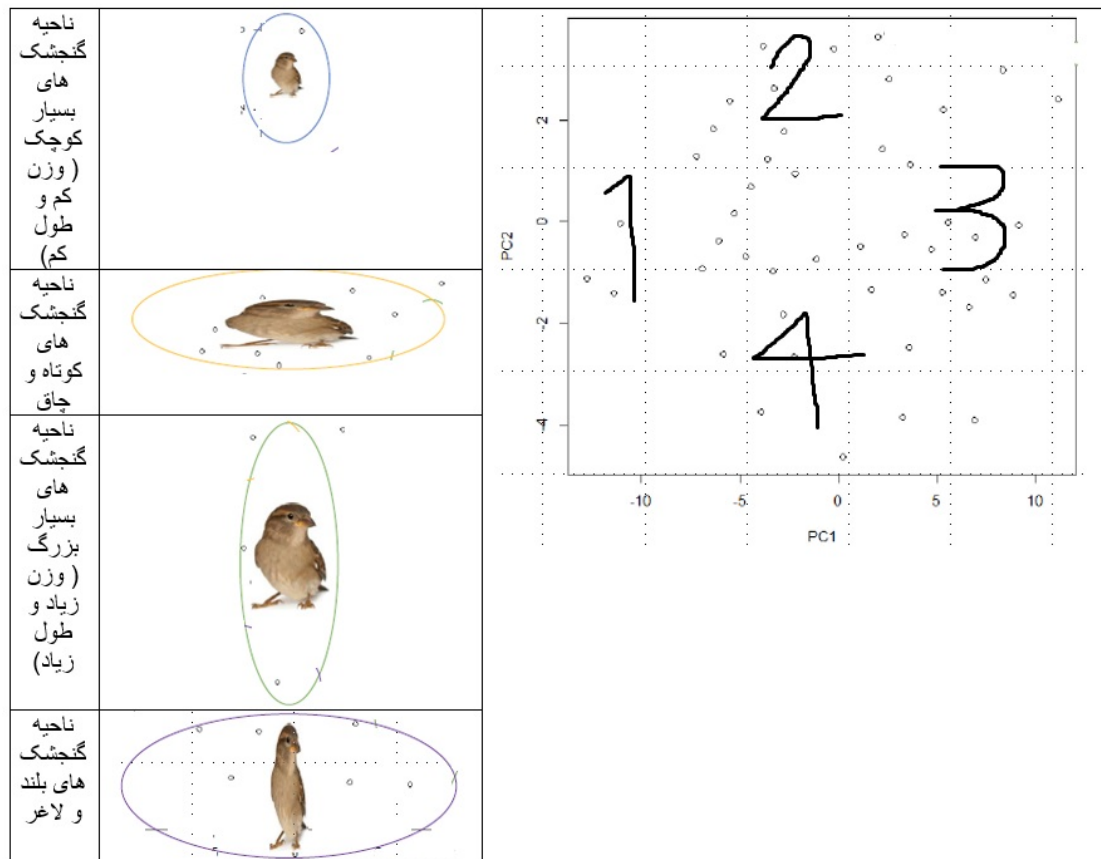
شکل ۲

۴. (۲۰ نمره) قابل توجه است که ۰/۵ درصد از افراد جامعه دارای یک بیماری خاص هستند. یک تست با احتمال ۰/۹۵ به درستی تایید می‌کند یک شخص این بیماری خاص را دارد. این تست برای شخصی که بیمار نیست با احتمال ۰/۹۶ و به درستی نتیجه منفی را گزارش می‌کند. حال آزمایش برای یک مراجعه کننده مثبت شده است؛ چقدر احتمال دارد این شخص این بیماری خاص را داشته باشد؟ بر اساس احتمالات گزارش شده بحث کنید که حساسیت و ویژگی برای این تست چه میزان هستند؟ همچنین شرح دهید کاربرد این معیارها در یادگیری ماشین برای چنین تست هایی چه مواردی می‌تواند باشد.

۵. (۲۰ نمره) اگر X_1 و X_2 به ترتیب طول و وزن تعدادی گنجشک باشند که اندازه های آنها ثبت شده است و ماتریس واریانس-کوواریانس آنها به صورت زیر محاسبه شده است:

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

ابتدا مولفه اصلی اول PC_1 و مولفه اصلی دوم PC_2 را که به صورت ترکیب خطی X_1 و X_2 هستند بر اساس مقادیر ویژه و برداره ویژه نرمال شده این ماتریس بدست آورید. هرکدام چه درصدی از اطلاعات موجود در متغیرهای اصلی را بیان می کنند؟ همچنین هر کدام از گنجشک ها با شرایط گفته شده در شکل، در کدام ناحیه (نواحی ۱ تا ۴) مختصات دوبعدی (PC_1, PC_2) قرار می گیرند؟



شکل ۳

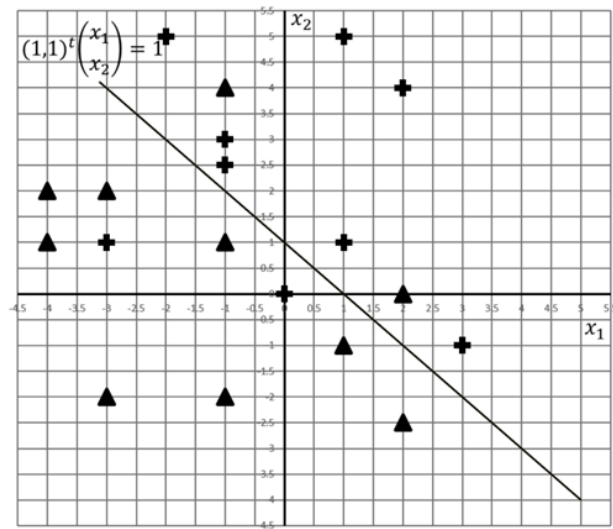
۶. (۲۰ نمره) فرض کنید بخواهیم با توجه به مجموعه داده

$$D_m = \{(z_1, y_1), (z_2, y_2), \dots, (z_m, y_m)\} \in \mathbb{R}^2 \times \{-1, 1\}$$

مسئله بهینه سازی به صورت زیر را برای یک SVM به منظور بدست آوردن پارامترهای w و b با وجود متغیر ξ_i حل کنیم:

$$\begin{aligned} \min_{w, b, \xi} \quad & \|w\|_2^2 + \frac{C}{m} \sum_{i=1}^m \xi_i \\ \text{s.t.} \quad & y_i(w^T z_i - b) \geq 1 - \xi_i \\ & \xi_i \geq 0 \quad \forall i \in \{1, 2, \dots, m\} \end{aligned}$$

برای $D_{1..}$ و پارامتر $C \geq 0$ بدست می‌آید: $\hat{w} = [1, 1]^T$ و $\hat{b} = 1$. اگر تعریف کنیم $\hat{f}(z) = \hat{w}^T z - \hat{b}$ آنگاه نمودار زیر مجموعه‌ای از نقاط $D_{1..}$ را نمایش می‌دهد. همچنین برای نمونه‌هایی از $D_{1..}$ که اینجا نمایش داده نشده‌اند داریم $y_i \hat{f}(z_i) > 1$ طوری که برچسب + برای کلاس $y = +1$ و برچسب - برای کلاس $y = -1$ است.



شکل ۴

(الف) با توجه به نمودار بالا خطوطی را رسم کنید که بیانگر باند موردنظر برای طبقه‌بندی کننده $\hat{w}^T z = \hat{b}$ باشد. قابل توجه است خطوطی که باندها را مشخص می‌کنند به صورت $\{z \in \mathbb{R}^2 : \hat{f}(z) = \pm 1\}$ هستند.

(ب) فرض کنید $\xi_{(x_1, x_2)}$ متغیر slack برای $z = (x_1, x_2)$ باشد. در آن صورت برای هر یک از گزاره‌های زیر کدام یک از علائم $>, <, =$ را مناسب می‌دانید؟

- $\xi_{(2, 4)} \quad \square \quad \xi_{(2, 0)} \quad \bullet$
- $\xi_{(-1, 1)} \quad \square \quad \xi_{(-1, -2)} \quad \bullet$
- $\xi_{(-3, 1)} \quad \square \quad \xi_{(-1, 2.5)} \quad \bullet$
- $\xi_{(2, 4)} \quad \square \quad \xi_{(-1, -2)} \quad \bullet$

(ج) قابل توجه است که می‌توان \hat{w} را به صورت $\hat{w} = \sum_{i=1}^m \alpha_i z_i$ نوشت که در آن صورت هر z_i با $\alpha_i \neq 0$ یک بردار پشتیبان نامیده می‌شود. برای هر نمونه در فضای آموزشی یکی از شرایط زیر وجود می‌آید:

۱. نمونه مورد نظر قطعا یک بردار پشتیبان است.
۲. نمونه مورد نظر قطعا یک بردار پشتیبان نیست.
۳. نمی‌توان با توجه به شرایط موجود گفت که یک نمونه می‌تواند یک بردار پشتیبان باشد یا خیر.

برای هر مدام از نمونه‌های زیر تعیین کنید کدام شرط از شرایط ۱ تا ۳ بالا برای آن محتمل‌تر است.

- $z = (2, 4) \quad \bullet$
- $z = (1, 1) \quad \bullet$
- $z = (2, 0) \quad \bullet$