## سؤال ۱ (۱۰ نمره)

۱.۱ در جبر خطی، روش Singular Value Decomposition روشی است که با استفاده از آن، یک ماتریس حقیقی $X$ به صورت زیر تجزیه می‌شود:

$$X = USV^T$$

اگر ماتریس $X$، $m \times n$ بعدی باشد که در آن بدون از دست دادن عمومیت $m \leq n$ است، آنگاه $U$ یک ماتریس $m \times n$، $S$ یک ماتریس $n \times n$ قطری، و $V^T$ نیز یک ماتریس $n \times n$ است. بعلاوه، $U$ و $V$ ماتریس‌های یکانی هستند ($UU^T = I$ و بطور مشابه برای $V$).
مجموعه‌ای از $N$ داده‌ی مشاهده‌شده‌ی $x^{(1)}$ تا $x^{(N)}$، را در نظر بگیرید و فرض کنید میانگین داده‌ها صفر است. می‌دانیم که در روش PCA ماتریس کوواریانس محاسبه می‌شود و بردارهای ویژه، راستاهای اصلی (Principal Components) هستند.

- الف) (۵ نمره) نشان دهید که اگر $USV^T$ تجزیه SVD ماتریس $X$ باشد، راستاهای اصلی در روش PCA برابر با ستون‌های $U$ هستند.
- ب) (۵ نمره) وقتی تعداد ابعاد بسیار بیشتر از تعداد داده‌هاست، بهتر است برای انجام PCA از SVD استفاده کنیم یا از محاسبه‌ی مستقیم بردارهای ویژه‌ی ماتریس کوواریانس؟

راهنمایی: بهترین ترتیبی زمانی ممکن برای محاسبه بردارهای ویژه ماتریس کوواریانس و SVD را با ذکر دلیل بیان و مقایسه کنید.

## سوال ۲ (۲۰ نمره)

۱. (۵ نمره) برای توابع Boolean زیر درخت تصمیم بهینه را طراحی کنید.

(آ)

$$(X_1 \wedge X_2) \vee X_3$$

(ب)

$$(X_1 \wedge X_2) \text{xor} (\neg X_1 \wedge X_2)$$

۲. (۴ نمره) در مبحث درخت تصمیم، برای انتخاب بهترین خصوصیتی که درخت را براساس آن تکه کنیم، از Information Gain استفاده نمودیم. معیار KL Divergence یک مفهوم مهم در شاخه تئوری اطلاعات است که به شرح زیر تعریف می‌شود.

$$KL(p \parallel q) = -\sum p(x) \log \frac{q(x)}{p(x)}$$

ثابت کنید معیار Information Gain برابر است با:

$$KL(p(x,y) \parallel p(x)q(x))$$

۳. (۶ نمره) مجموعه داده‌ای در جدول ۱ در ارتباط با سرطان و ویژگی‌های موثر بر آن در اختیار داریم. از روی این داده‌ها یک درخت تصمیم به صورت زیر پیشنهاد شده است. چنانچه از روش ID3 استفاده شود آیا باز هم ویژگی سن در ریشه قرار می‌گیرد (با محاسبات پاسخ دهید)؟

## سوال ۳ (۱۰ نمره)

می‌خواهیم مسائل غیرخطی زیر را با تغییر متغیر یا اعمال مشابه به مسائل رگرسیون خطی با تابع خطای MSE تبدیل کنیم. پارامترهای مدل $\alpha_i$ ها هستند. با ذکر دلیل مشخص کنید که آیا این کار ممکن است یا خیر.

۱.

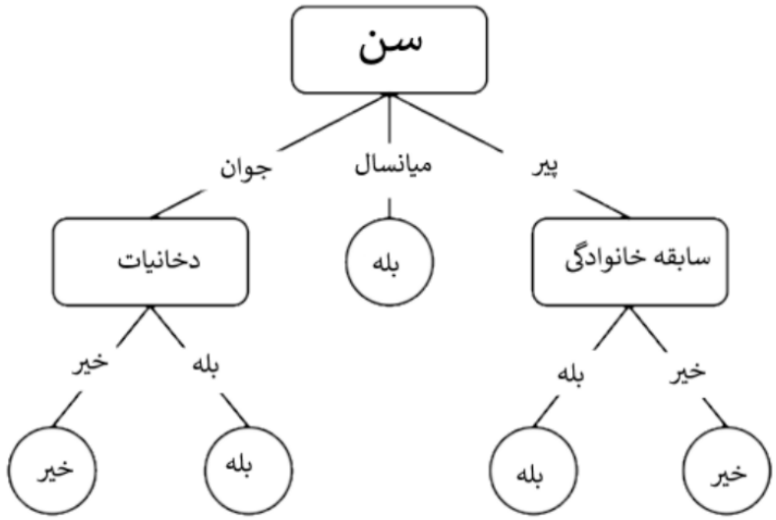
$$y = \alpha x_1^2 x_2^3 + \epsilon$$

۲.

$$y = x_1^{\alpha_1} x_2^{\alpha_2} + \epsilon$$

۳.

$$y = \log(x_1^{\alpha_1} x_2^{\alpha_2}) + \epsilon$$




شکل ۱: ساختار درخت تصمیم

جدول ۱: داده‌های سرطان

| سن | فعالیت بدنی | دخانیات | سابقه خانوادگی سرطان | سرطان |
|---|---|---|---|---|
| جوان | زیاد | خیر | بله | خیر |
| جوان | زیاد | خیر | خیر | خیر |
| میانسال | زیاد | خیر | بله | بله |
| پیر | متوسط | خیر | بله | بله |
| پیر | کم | بله | بله | بله |
| پیر | کم | بله | خیر | خیر |
| میانسال | کم | بله | خیر | بله |
| جوان | متوسط | خیر | بله | خیر |
| جوان | کم | بله | خیر | بله |
| پیر | متوسط | بله | بله | بله |
| جوان | متوسط | بله | خیر | بله |
| میانسال | متوسط | خیر | خیر | بله |
| میانسال | زیاد | بله | بله | بله |
| پیر | متوسط | خیر | خیر | خیر |

## سوال ۴ (۱۰ نمره)

توجه کنید که به دست آوردن رابطه نهایی کافی نیست و چگونگی رسیدن به روابط نهایی و نحوه بهینه‌سازی بایستی نوشته شود.

۱. فرض کنید متغیر تصادفی $X$ از توزیع نرمال با میانگین $\mu$ و واریانس $\sigma^2$ آمده است. همچنین $n$ نمونه $x_1$ الی $x_n$ مجموعه مشاهدات $(D)$ را تشکیل می‌دهند.

(آ) (۴ نمره) با استفاده از رویکرد MLE پارامترهای $\mu$ و $\sigma^2$ را تخمین بزنید.

(ب) (۶ نمره) با فرض معلوم بودن پارامتر $\sigma^2$ و دانش پیش‌روی پارامتر $\mu$ که توزیع نرمال با پارامتر $\mu_0$ و $\sigma_0$ است، ابتدا توزیع پسین $\mu$ را با توجه به مشاهدات ($P(\mu|D, \mu_0, \sigma_0)$) محاسبه کنید و سپس با رویکرد MAP مقدار $\mu$ را تخمین بزنید.

تابع توزیع نرمال به صورت زیر است:

$$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$