

توجه: استفاده از کتاب، جزوه، اسلایدهای درس، اینترنت و مشورت در حین امتحان غیر مجاز است و تقلب محسوب می‌شود. در صورت تشخیص تقلب، نمره کل امتحان صفر منظور خواهد شد.

توجه: امتحان از ۱۱۰ نمره است و برای کامل شدن باید ۱۰۰ نمره کسب شود.

سوال ۱ سوالات پاسخ کوتاه (۲۴ نمره)

در هر یک از موارد زیر درست یا غلط بودن آن را مشخص کنید و حداکثر در ۲ جمله به صورت مختصر علت را توضیح دهید. (هر مورد ۴ نمره)

(الف) معمولاً مدل‌های غیرپارامتریک مفروضات کمتری نسبت به مدل‌های پارامتریک در نظر می‌گیرند.

(ب) اگر در یک مسئله رگرسیون رابطه‌ی بین $x = [x_1 \ x_2 \ x_3]^T$ و y را به صورت $y = \alpha_1 x_1^2 x_2 + \alpha_2 x_3$ در نظر بگیریم. می‌توان با استفاده از رگرسیون خطی، مقادیر پارامترهای α_1, α_2 را بدست آورد.

(ج) خطای طبقه‌بند نزدیک‌ترین همسایه (1-NN) روی داده‌های آموزشی (در صورت سازگاری داده‌ها) همواره صفر است.

(د) اگر در یک مسئله رگرسیون رابطه‌ی بین $x = [x_1 \ x_2 \ x_3]^T$ و y را به صورت $y = (x_1 x_2)^{\alpha_1} + x_3^{\alpha_2}$ در نظر بگیریم. می‌توان با استفاده از رگرسیون خطی، مقادیر پارامترهای α_1, α_2 را بدست آورد.

(ه) اضافه کردن منظم‌ساز نرم ۱ به مجموع مجذور خطا (SSE) در یک مسئله رگرسیون خطی $y = w^T x$ ، معمولاً باعث تنک (sparse) شدن بردار w می‌شود. با رسم شکل توضیح دهید.

(و) از روش پنجره پارزن نمی‌توان برای طبقه‌بندی استفاده کرد.

سوال ۲ رگرسیون خطی (۱۶ نمره)

در یک مسئله رگرسیون خطی، مجموعه داده‌ی $D = \{(x_1, y_1), \dots, (x_n, y_n)\}$ را در اختیار داریم. رابطه‌ی احتمالاتی میان x و y را به صورت زیر در نظر می‌گیریم:

$$y_i = wx_i + \epsilon_i$$

$$\epsilon_i \sim \mathcal{N}(0, 1)$$

که در آن w پارامتر مدل و ϵ_i یک نویز گاوسی با میانگین صفر و واریانس ۱ است.

(الف) با فرض i.i.d بودن داده‌ها، تابع log-likelihood را تشکیل دهید و نشان دهید که بیشینه کردن تابع log-likelihood روی پارامتر w معادل است با کمینه کردن مجموع مجذور خطا. به عبارت دیگر نشان دهید:

$$\arg \max_w \log P(D|w) = \arg \min_w \sum_{i=1}^n (y_i - wx_i)^2$$

(ب) اگر بخواهیم از دیدگاه بیز به این مسئله نگاه کنیم، باید پارامتر w را یک متغیر تصادفی در نظر بگیریم و برای آن یک توزیع احتمال پیشین داشته باشیم. حال فرض کنید توزیع احتمال پیشین زیر را برای w داشته باشیم:

$$P(w) = \mathcal{N}(0, \sigma^2)$$

نشان دهید که تخمین گر MAP برای w معادل است با کمینه کردن SSE با یک ترم منظم ساز درجه ۲:

$$\arg \max_w \log P(w|D) = \arg \min_w \sum_{i=1}^n (y_i - wx_i)^2 + \lambda w^2$$

ضریب λ چه رابطه‌ای با σ^2 دارد؟

سوال ۳ طبقه بند بیز ساده (۱۰ نمره)

مجموعه داده‌ی زیر را در نظر بگیرید که در آن متغیر y برچسب است و متغیرهای A, B, C ویژگی‌های باینری هستند. طبقه‌بند بیز ساده (Naïve Bayes)، نمونه‌های $(0,0,1)$ و $(1,1,1)$ را به چگونه دسته‌بندی می‌کند؟ (در حالت تساوی احتمال، برچسب صفر ارجح است).

A	B	C	y
0	0	1	0
0	1	0	0
1	1	0	0
0	0	1	1
1	1	1	1
1	0	0	1
1	1	0	1

سوال ۴ Discriminant function (۱۵ نمره)

یک طبقه بند بیز که در آن احتمال پیشین دو کلاس برابر است را در نظر بگیرید. احتمال دو کلاس به صورت زیر است:

$$P(x|y=0) = \frac{2}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad x \geq 0$$

$$P(x|y=1) = e^{-x} \quad x \geq 0$$

بازه‌ی ناحیه‌ی کلاس اول و دوم را بدست آورید. ($\ln \frac{2}{\sqrt{2\pi}} \cong -0.1$)

سوال ۵ درخت تصمیم (۲۰ نمره)

می‌خواهیم با استفاده از داده‌های زیر، یک درخت تصمیم برای دسته‌بندی دانشجویان به دو دسته‌ی تنبل (L) و زنگ (D) استفاده کنیم.

Feature1	Feature2	Feature3	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

مقادیر زیر می‌تواند در محاسبات به شما کمک کند:

$$\log_2 0.1 = -3.32, \quad \log_2 0.2 = -2.32, \quad \log_2 0.3 = -1.73, \quad \log_2 0.4 = -1.32, \quad \log_2 0.5 = -1$$

توجه: حق استفاده از ماشین حساب را ندارید!

الف) (۴ نمره) آنتروپی شرطی $H(feature2|feature1 = N)$ را محاسبه کنید.

ب) (۸ نمره) در الگوریتم *ID3*، کدام ویژگی به عنوان ریشه‌ی درخت انتخاب می‌شود؟

ج) (۵ نمره) درخت تصمیم کامل را بدست آورید.

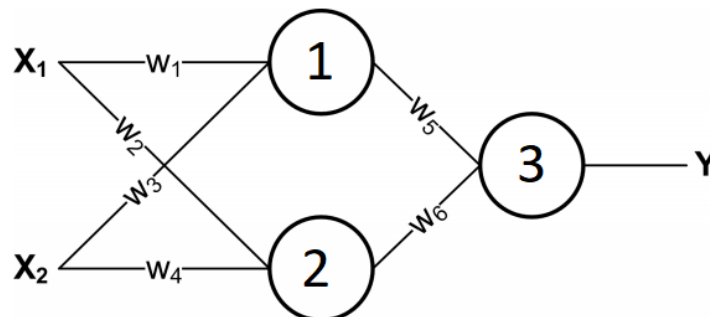
د) (۳ نمره) مقادیر *recall* و *accuracy* این درخت تصمیم روی داده‌های آموزشی چقدر است؟

سوال ۶ شبکه‌های عصبی (۱۵ نمره)

شبکه عصبی زیر را در نظر بگیرید. در نورون‌هایی که با شماره‌های ۱ تا ۳ مشخص شده‌اند باید تابع فعالیت (Activation Function) قرار گیرد. می‌خواهیم از دو تابع فعالیت زیر استفاده کنیم:

- S: signed sigmoid function $S(a) = \text{sign}(\sigma(a) - 0.5) = \text{sign}\left(\frac{1}{1+e^{-a}} - 0.5\right)$
- L: Linear function $L(a) = c a$

که در روابط بالا $a = \sum_i w_i X_i$ است.



الف) (۷ نمره) اگر بخواهیم شبکه‌ی بالا معادل یک رگرسیون خطی عمل کند، یعنی $Y = \beta_1 X_1 + \beta_2 X_2$ ، تابع فعالیت مناسب (S یا L) برای هر یک از نورون‌های ۱ تا ۳ چه خواهد بود؟

ب) (۸ نمره) اگر بخواهیم شبکه‌ی بالا معادل یک طبقه‌بند لاجستیک رگرشن عمل کند، یعنی:

$$P(Y = 1|X) = \frac{1}{1 + \exp(-\beta_1 X_1 - \beta_2 X_2)}$$

تابع فعالیت مناسب برای نورون‌های ۱ تا ۳ چیست؟

سوال ۷ (۱۰ نمره)

مصالحه‌ی بایاس-واریانس را توضیح دهید.