

**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 (۲۰ 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) همواره  $H(X|Y=y) \leq H(X)$  صادق است.

ب) طبقه‌بند بهینه بیز، بهترین طبقه‌بندی است که می‌توان برای جداسازی یک مسئله دو کلاسه طراحی کرد.

ج) اگرچه در روش‌های استنباط آماری کلاسیک تلاش بر طراحی تخمیک‌های ناریب (unbiased) بوده است، در یادگیری ماشین، مقداری از خطای بایاس پذیرفته می‌شود.

د) استفاده از رویکرد بیز برای تخمین پارامترهای توزیع می‌تواند مانع از بیش برازش (overfitting) شود.

ه) استفاده از معیار Information Gain برای ساخت درخت در شرایطی که بعضی از ویژگی‌ها (features) دارای حالت‌های زیادی هستند، مناسب نیست.

### سوال ۲ تخمین بیز (۱۶ نمره)

فرض کنید متغیر تصادفی  $X$  دارای توزیع احتمال  $N(\mu, 1)$  باشد. اگر توزیع احتمال پیشین روی پارامتر  $\mu$  را یک توزیع ریلی به صورت زیر در نظر بگیریم:

$$P(\mu) = \frac{\mu}{\sigma^2} \exp\left(-\frac{\mu^2}{2\sigma^2}\right) \quad \mu \geq 0$$

الف) (۱۲ نمره) اگر مجموعه داده‌ی  $D = \{x_1, \dots, x_n\}$  را در اختیار داشته باشیم. نشان دهید که تخمین گر  $\hat{\mu}_{MAP}$  به صورت زیر بدست می‌آید:

$$\hat{\mu}_{MAP} = \frac{Z}{2R} \left( 1 + \sqrt{1 + \frac{4R}{Z^2}} \right)$$

که در رابطه‌ی بالا داریم:

$$Z = \sum_{i=1}^n x_i \quad R = n + \frac{1}{\sigma^2}$$

ب) (۴ نمره) آیا توزیع احتمال فوق، یک conjugate prior برای پارامتر  $\mu$  است؟ توضیح دهید.

### سوال ۳ Bayesian Decision (۱۵ نمره)

یک مسأله طبقه‌بندی دو کلاسه با احتمال پیشین مساوی را در نظر بگیرید. فرض کنید داده‌های دو کلاس بر اساس توزیع‌های زیر تولید می‌شوند:

$$p(x|y=1) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right) \quad x \geq 0$$

$$P(x|y=2) = \theta x \exp(-\theta x) \quad x \geq 0$$

که در رابطه‌ی بالا  $\sigma > 0$  و  $\theta > 0$  پارامترهای مدل هستند. ناحیه‌ی مربوط به دو کلاس را در طبقه‌بند بیز بدست بیاورید.

### سوال ۴ رگرسیون خطی (۸ نمره)

فرض کنید که در فضای ویژگی یک بعدی ( $x$ ) و یک متغیر خروجی پیوسته ( $y$ ) درصدد یافتن مدل برای داده‌ها هستیم. می‌دانیم رابطه زیر برقرار است

$$y \sim N(\exp(wx), 1)$$

که در آن  $N$  توزیع نرمال با میانگین  $\exp(wx)$  و واریانس یک است. شما از کدام روش‌های بیان شده در درس برای یافتن این مدل استفاده می‌کنید؟ پاسخ خود را شرح دهید.

### سوال ۵ رگرسیون خطی (۱۵ نمره)

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

$$y_i = w^T x_i + \epsilon_i$$

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

که در آن  $w \in \mathbb{R}^d$  پارامتر مدل و  $\epsilon_i$  یک نویز گاوسی با میانگین صفر و واریانس ۱ است.

اگر بخواهیم از دیدگاه بیز به این مسأله 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 لاپلاس را برای  $w$  داشته باشیم:

$$P(w) = \frac{1}{(2b)^d} \exp\left(-\frac{|w|}{b}\right)$$

که در رابطه‌ی بالا،  $|w|$  نرم ۱ بردار  $w$  است. نشان دهید که تخمین گر MAP برای  $w$  معادل است با کمینه کردن SSE با یک ترم منظم‌ساز نرم ۱:

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

رابطه‌ی بین  $\lambda$  و  $b$  چیست؟

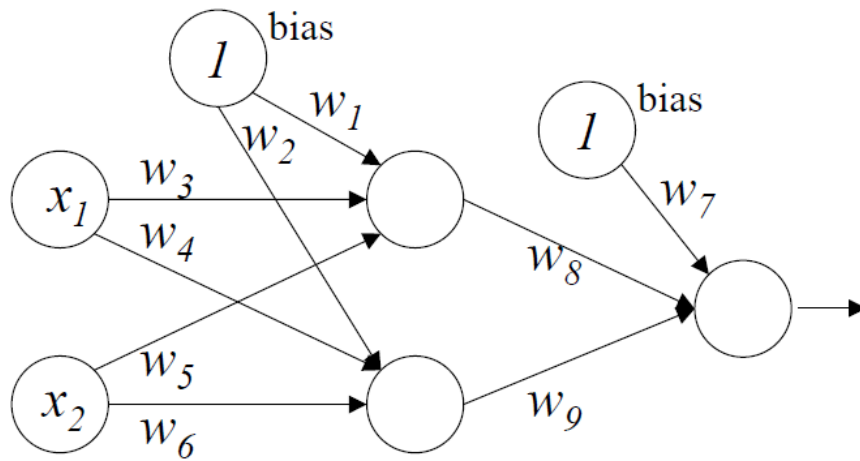
### سوال ۶ درخت تصمیم (۱۴ نمره)

فرض کنید که ویژگی باینری  $A$ ،  $B$  و  $C$  به همراه برچسب کلاس هر داده بیان شده باشد (جدول زیر). می‌خواهیم درخت تصمیم با کمترین عمق را برای داده‌های داده شده بیابیم. آیا الگوریتم ID3 (بدون هرس کردن) درخت بهینه را می‌یابد؟ با رسم درخت‌های مربوطه جواب خود را شرح دهید.

A	B	C	Class
1	1	0	0
1	0	1	1
0	1	1	1
0	0	1	0

### سوال ۷ شبکه عصبی (۱۲ نمره)

شبکه عصبی زیر را برای طبقه بندی دو کلاسه در نظر بگیرید. فرض کنید که لایه های میانی از تابع فعالسازی خطی  $h(z) = cz$  و تابع سیگموئید  $g(z) = \frac{1}{1+e^{-z}}$  در لایه خروجی استفاده می کند. این شبکه در صدد یادگیری  $P(Y = 1|X, w)$  است که در آن  $X = (x_1, x_2)$  و  $W = (w_1, w_2, \dots, w_9)$  است.



الف) (۸ نمره) خروجی شبکه عصبی  $P(Y = 1|X, w)$  را بر حسب پارامترهای شبکه  $(W, X)$  و ثابت  $C$  نوشته و مرز تصمیم نهایی را به دست آورید.

ب) (۴ نمره) آیا می توان شبکه عصبی بدون لایه مخفی به دست آورد که معادل شبکه عصبی فوق باشد؟ در صورت وجود شبکه پیشنهادی را رسم کنید.

### سوال ۸ K-NN (۱۰ نمره)

در چه شرایطی بازدهی طبقه بندی نزدیکترین همسایه با طبقه بندی بیز یکسان است. در مورد شروط تحقق بازدهی بهینه طبقه بندی نزدیکترین همسایه بحث کنید؟