

توجه: استفاده از کتاب، جزوه، اسلایدهای درس، اینترنت و مشورت در حین امتحان غیر مجاز است و تقلب محسوب می‌شود.

نمره امتحان از ۱۱۰ نمره است و برای کامل شدن باید ۱۰۰ امتیاز کسب کنید.

### سوال ۱ سوالات پاسخ کوتاه (۲۸ 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)

- با افزایش داده‌های آموزشی می‌توان خطای بایاس را کاهش داد.
- با اضافه کردن منظم‌ساز<sup>۱</sup> به تابع هزینه، خطای بایاس افزایش و خطای واریانس کاهش می‌یابد.
- با استفاده از الگوریتم ID3 برای ساخت درخت تصمیم، درخت بهینه (کم ارتفاع‌ترین درخت) بدست می‌آید.
- تفاضل دو تابع kernel همواره یک کرنل معتبر است.
- اگر در یک مسأله رگرسیون رابطه‌ی بین  $x = [x_1 \ x_2 \ x_3]^T$  و  $y$  را به صورت  $y = \alpha_1 x_1^2 + \alpha_2 \sin x_2 x_3$  در نظر بگیریم. می‌توان با استفاده از رگرسیون خطی، مقادیر پارامترهای  $\alpha_1, \alpha_2$  را بدست آورد.
- روش پنجره پارزن زمانی که تعداد نمونه‌ها بینهایت باشد، همواره به توزیع اصلی داده‌ها میل می‌کند.
- روش SVM نسبت به عدم توازن در تعداد نمونه‌های دو کلاس مقاوم است.

### سوال ۲ (۳۰ نمره)

متغیر تصادفی  $X$  را با توزیع احتمال نمایی با پارامتر  $\lambda$  را در نظر بگیرید:

$$P(X) = \lambda e^{-\lambda x}$$

فرض کنید یک مجموعه داده‌ای شامل  $n$  نمونه‌ی  $D = \{X_1, \dots, X_n\}$  از این متغیر تصادفی را در اختیار داریم.

الف) (۱۰ نمره) تابع لگاریتم درست‌نمایی<sup>۲</sup> را تشکیل داده و تخمین‌گر بیشینه درست‌نمایی<sup>۳</sup> را برای پارامتر  $\lambda$  بدست آورید.

ب) (۱۰ نمره) توزیع احتمال پیشین<sup>۴</sup> زیر را برای پارامتر  $\lambda$  در نظر بگیرید:

$$p(\lambda) = \text{Gamma}(\lambda | \alpha, \beta) = c \lambda^{\alpha-1} e^{-\beta\lambda}$$

که در رابطه‌ی بالا،  $c$  یک ضریب ثابت است و  $\alpha$  و  $\beta$  پارامترهای توزیع گاما هستند. توزیع احتمال پسین<sup>۵</sup> را برای پارامتر  $\lambda$  بدست آورید (نیازی به محاسبه ضریب ثابت توزیع پسین نیست):

$$p(\lambda|D) = ?$$

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

<sup>1</sup> Regularization Term

<sup>2</sup> Log-Likelihood

<sup>3</sup> Maximum Likelihood

<sup>4</sup> Prior distribution

<sup>5</sup> Posterior Distribution

د) (۳ نمره) با استفاده از توزیع احتمال پیشین فوق، تخمین گر MAP برای پارامتر  $\lambda$  چیست؟ (راهنمایی: مقدار بیشینه توزیع گاما در نقطه  $\lambda = \frac{a-1}{\beta}$  رخ می دهد).

ه) (۳ نمره) آیا اگر  $n \rightarrow +\infty$  آنگاه، تخمین گر MAP به تخمین گر ML میل می کند؟

### سوال ۳ (۱۵ نمره)

فرض کنید برای  $x \in [0, 3]$  توزیع های مربوط به دو کلاس به صورت زیر باشد:

$$p(x|y=0) = \frac{x^2}{9}$$

$$p(x|y=1) = \frac{x}{9} + \frac{1}{6}$$

بازه های نواحی کلاس اول و دوم را توسط طبقه بند بهینه بیز بدست آورید (احتمال پیشین کلاس ها را یکسان در نظر بگیرید).

### سوال ۴ (۲۲ نمره)

الف) (۱۰ نمره) ثابت کنید در طبقه بند  $SVM$ ، اندازه ی حاشیه برابر است با  $\frac{2}{\|w\|_2}$  که در آن  $w$  بردار نرمال ابرصفحه ی جدا کننده است.

ب) (۶ نمره) توضیح دهید که در طبقه بند  $SVM$ ، مقدار پارامتر عرض از مبدا چگونه محاسبه می شود.

ج) (۶ نمره) در روش  $kernel-SVM$ ، توضیح دهید که یک نمونه ی جدید چگونه دسته بندی می شود و چرا برای طبقه بندی آن، نیازی به داشتن همه ی داده های آموزشی نیست؟

### سوال ۵ (۱۵ نمره)

یک مسأله رگرسیون خطی با مجموعه داده ی آموزشی  $\{(x_1, y_1), \dots, (x_n, y_n)\}$  را در نظر بگیرید ( $y_i \in \mathbb{R}$  و  $x_i \in \mathbb{R}^d$ ). اگر از تابع هزینه زیر استفاده کنیم:

$$L(w) = \sum_{i=1}^n (w^T x_i - y_i)^2 + \lambda \|w\|_2^2$$

که در آن  $\lambda$  یک ضریب ثابت مثبت است، فرم بسته مقدار بهینه  $w$  را بدست آورید.