

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.

سوال ۱ سوالات پاسخ کوتاه (۲۱ 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 ۳ نمره)

- احتمال بیش‌برازش^۱ برای طبقه‌بندی که روی داده‌های کمتری آموزش دیده باشد، کمتر است.
- مدل‌های ساده‌تر دارای خطای بایاس بیشتر و خطای واریانس کمتری نسبت به مدل‌های پیچیده‌تر هستند.
- زمانی که داده‌های آموزشی کم باشد، استفاده از تخمین‌گرها بیزی بهتر از تخمین‌گر بیشینه‌درست‌نمایی^۲ است.
- اگر اطلاعات متقابل^۳ دو متغیر تصادفی صفر باشد، آن‌گاه همبستگی^۴ آن‌ها نیز صفر است.
- استفاده از ترم‌های منظم‌ساز^۵ باعث کاهش خطای واریانس می‌شود.
- اگر در یک طبقه‌بند بیز ساده^۶ احتمال هر ویژگی به شرط برچسب را یک توزیع گاوسی به صورت زیر در نظر بگیریم:

$$P(x_i | y = 0) = N(\mu_i, 1) \quad i = 1, \dots, d$$

$$P(x_i | y = 1) = N(\alpha_i, 1) \quad i = 1, \dots, d$$

- که در آن x_i بعد i -ام بردار ورودی است. آن‌گاه مرز جدا کننده دو کلاس خطی خواهد شد.
- اگر داده‌ها به صورت خطی جدایی پذیر باشند، الگوریتم پرسپترون بسته به اینکه مقدار اولیه وزن‌ها را چگونه انتخاب کنیم، ممکن است نتواند داده‌های دو کلاس را از هم تفکیک کند.

سوال ۲ (۲۹ نمره)

متغیر تصادفی X را با توزیع احتمال پواسون با پارامتر λ را در نظر بگیرید:

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

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

^۱ overfit

^۲ Maximum Likelihood

^۳ Mutual Information

^۴ Correlation

^۵ Regularizer

^۶ Naïve Bayes

الف) (۱۰ نمره) تابع لگاریتم درستنمایی^۱ را تشکیل داده و تخمین گر بیشینه درستنمایی^۲ را برای پارامتر λ بدست آورید.

ب) (۱۰ نمره) توزیع احتمال پیشین^۳ زیر را برای پارامتر λ 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 α و β پارامترهای توزیع گاما هستند. توزیع احتمال پسین^۴ را برای پارامتر λ بدست آورید (نیازی به محاسبه ضریب ثابت توزیع پسین نیست):

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

ج) (۳ نمره) آیا توزیع احتمال پیشین فوق، یک conjugate prior برای پارامتر λ است؟ توضیح دهید.

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

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

سوال ۳ (۲۰ نمره)

در یک مسأله رگرسیون، می‌خواهیم رابطه‌ی بین ورودی و مقدار خروجی را به صورت زیر مدل کنیم:

$$y = \exp wx$$

در رابطه‌ی بالا، $y \in \mathbb{R}$ و $x \in \mathbb{R}$ است و $w \in \mathbb{R}$ پارامتر مدل است. فرض کنید مجموعه داده‌ی آموزشی $D = \{(x_1, y_1), \dots, (x_n, y_n)\}$ را در اختیار داریم.

الف) (۵ نمره) تابع هزینه مجموع مجذور خطا^۵ را برای مجموعه داده‌ی D تشکیل دهید.

ب) (۵ نمره) اگر بخواهیم با استفاده از روش کاهش گرادیان^۶ مقدار بهینه w را بدست آوریم، رابطه‌ی بروز رسانی w چه خواهد بود؟ به عبارت دیگر، w_{t+1} چگونه از w_t بدست می‌آید.

ج) (۱۰ نمره) با انجام محاسبات نشان دهید که برای کمینه کردن تابع هزینه، مقدار بهینه پارامتر w باید در کدامیک از روابط زیر صدق کند؟

- $\sum_{i=1}^n x_i \exp wx_i = \sum_{i=1}^n x_i y_i \exp wx_i$
- $\sum_{i=1}^n x_i^2 \exp wx_i = \sum_{i=1}^n x_i y_i \exp wx_i$
- $\sum_{i=1}^n x_i y_i = \sum_{i=1}^n x_i \exp wx_i$
- $\sum_{i=1}^n x_i \exp 2wx_i = \sum_{i=1}^n x_i y_i \exp wx_i$
- $\sum_{i=1}^n \exp wx_i = \sum_{i=1}^n x_i y_i \exp wx_i$

^۱ Log-Likelihood

^۲ Maximum Likelihood

^۳ Prior distribution

^۴ Posterior Distribution

^۵ Sum of squared error (SSE)

^۶ Gradient descent

سوال ۴ (۱۸ نمره)

در این مسأله می‌خواهیم سلیقه علی را برای غذاهای مختلف تشخیص دهیم. برای این کار می‌خواهیم از یک درخت تصمیم استفاده کنیم. فرض کنید داده‌های زیر را در اختیار داریم:

نظر علی	دمای غذا	طعم غذا	اندازه غذا
بد	گرم	ترش	کوچک
خوب	سرد	ترش	بزرگ
بد	گرم	شیرین	کوچک
خوب	سرد	ترش	کوچک
بد	گرم	شور	کوچک
بد	گرم	شیرین	بزرگ
خوب	گرم	شور	بزرگ
خوب	سرد	ترش	بزرگ
بد	سرد	شور	کوچک
بد	گرم	شور	کوچک

الف) (۵ نمره) آنتروپی نظر علی چقدر است؟

ب) (۵ نمره) میزان information gain برای ویژگی «طعم غذا» چقدر است؟

ج) (۸ نمره) اگر یک درخت تصمیم با ارتفاع ۱ با استفاده از ویژگی «طعم غذا» بسازیم، مقادیر specificity، recall، precision و accuracy برای این طبقه‌بند روی داده‌های آموزشی را بدست آورید.

سوال ۵ (۱۵ نمره)

فرض کنید برای $x \in [0, 3]$ توزیع‌های مربوط به دو کلاس و ماتریس هزینه λ را به صورت زیر تعریف کرده باشیم

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

$$p(x|y=1) = \frac{2x}{9}$$

$$\lambda = \begin{bmatrix} 1 & 5 \\ 7 & 1 \end{bmatrix}$$

بازه‌های نواحی کلاس اول و دوم را توسط طبقه‌بند بهینه بیز بدست آورید (احتمال پیشین کلاس‌ها را یکسان در نظر بگیرید).

سوال ۶ (۱۲ نمره)

الف) (۶ نمره) بزرگترین مشکل شبکه‌های عصبی در مرحله آموزش چیست؟ دو مورد از راهکارهای پیشنهادی برای رفع این مشکل را بیان کنید.

ب) (۶ نمره) در حالت حدی ($n \rightarrow \infty$) تحت چه شرایطی عملکرد روش KNN با طبقه‌بند بهینه بیز یکسان خواهد شد؟ توضیح دهید.