

**سوال ۱: Regression(20)**

ثابت کنید افزودن عبارت  $\lambda w^T w$  به تابع خطای رگرسیون معادل کمینه کردن تابع خطای MSE در یک کره  $d$  بعدی حول مبدأ است. ( $d$  برابر با ابعاد مسئله است).  
 راهنمایی: تابع خطا را به صورت

$$\operatorname{argmin}_w \text{MSE}(Y, w, X) + \lambda w^T w \quad (1)$$

در نظر بگیرید. می‌دانیم کمینه‌ی این تابع یکتاست.

پاسخ:

فرض می‌کنیم:

$$\hat{w} = \operatorname{argmin}_w \text{MSE}(w) + \lambda w^T w$$

سپس مقدار  $s = \hat{w}^T \hat{w}$  را تعریف می‌کنیم. کره  $R$  به صورت  $R = \{x | x^T x \leq s\}$  در نظر می‌گیریم و فرض می‌کنیم:

$$\tilde{w} = \operatorname{argmin}_{w \in R} \text{MSE}(w)$$

می‌دانیم که  $\hat{w} \in R$  پس:

$$\text{MSE}(\tilde{w}) \leq \text{MSE}(\hat{w})$$

هم چنین:

$$\begin{aligned} \text{MSE}(\hat{w}) + \lambda \hat{w}^T \hat{w} &\leq \text{MSE}(\tilde{w}) + \lambda \tilde{w}^T \tilde{w} \leq \text{MSE}(\hat{w}) + \lambda \tilde{w}^T \tilde{w} \\ &\leq \text{MSE}(\hat{w}) + \lambda s = \text{MSE}(\hat{w}) + \lambda \hat{w}^T \hat{w} \end{aligned}$$

در نتیجه  $\text{MSE}(\tilde{w}) + \lambda \tilde{w}^T \tilde{w} = \text{MSE}(\hat{w}) + \lambda \hat{w}^T \hat{w}$ . پس  $\tilde{w} = \hat{w}$ .

**سوال ۲: True/False(10)**

درستی یا نادرستی عبارتهای زیر را با ذکر دلیل (در حد یک جمله) مشخص کنید.

- روش‌های منظم‌سازی lasso و ridge برای انتخاب ویژگی‌ها نیز کاربرد دارند. (۵ نمره)
- تابع هزینه‌ی رگرسیون خطی را در حالتی که خطا بر حسب کمترین مربعات نوشته شده است در نظر بگیرید. با افزودن عبارت منظم‌سازی  $L_2$  به تابع هزینه نمی‌توانیم مقدار خطا را روی دادگان آموزش کاهش دهیم. (۵ نمره)

پاسخ:

- غلط، روش منظم‌سازی lasso به دلیل افزودن sparsity می‌تواند برای انتخاب ویژگی مورد استفاده قرار بگیرد. اما روش ridge تنها اندازه‌ی بردار وزن را محدود می‌کند و روشی برای انتخاب ویژگی نیست.
- درست، مقدار  $w$  بدست آمده در تابع هزینه بدون جمله‌ی منظم‌سازی کمینه است. بنابراین با افزودن جمله‌ی منظم‌سازی نمی‌توان به  $w$ ی رسید که خطا را روی دادگان آموزش کمتر کند.

در هر یک از موارد زیر، مشخص کنید عملکرد PCA بهتر است یا t-SNE؟

(آ) داده‌هایی که حول یک منحنی به شکل دایره پخش شده‌اند.

(ب) داده‌های خطی با نویز کم

(ج) مجموعه‌ای با تعداد زیادی داده

(د) داده‌هایی که feature های آن همبستگی بالایی دارند.

پاسخ

T-SNE ●

PCA ●

PCA ●

T-SNE ●