



دانشکده‌گان علوم  
دانشکده ریاضی، آمار و علوم کامپیوتر

مهلت تحویل: ۱۸ فروردین

تمرین عملی اول

**بخش ۱.** در این بخش شما با دو دیتاست تک‌بعدی مواجه می‌شوید و  $data\ fitting$  را با استفاده از روش  $least\ squares$  انجام می‌دهید. برای این منظور دو دیتاست با نام‌های  $data1.csv$  و  $data2.csv$  در اختیار شما قرار داده شده است. باید در این بخش پس از حذف نویز از دیتا به منظور هموارتر کردن تابع، یک چندجمله‌ای به درجه مناسب به هر دیتاست  $fit$  شود.

**بخش ۲.** عبارت‌های منظم‌ساز متنوعی با توجه به مسئله در فضای  $regularized\ least\ squares$  به کار می‌روند. یکی از معروف‌ترین‌های آن‌ها نورم ۲ بردار پارامترهاست که به  $regression$  به این روش،  $ridge\ regression$  گفته می‌شود. در مسئله پایه  $least\ squares$  داریم:

$$\min_x ||Ax - b||^2 \quad (1)$$

حال در  $ridge\ regression$  مسئله به فرم زیر تبدیل می‌شود که در آن ماتریس  $D$  همانی است.

$$\min_x ||Ax - b||^2 + \lambda ||Dx||^2 \quad (\lambda > 0) \quad (2)$$

پس از پیاده‌سازی  $ridge\ regression$  با دیتاست  $diabetes.csv$  (که میزان پیشرفت دیابت بیماران را اندازه‌گیری کرده است) به سوالات زیر پاسخ دهید.

- مقدار پارامتر  $\lambda$  در بردار  $x$  به دست آمده چه تاثیری دارد؟ مسئله را با مقادیر مختلف  $\lambda$  حل کنید.
- این عبارت منظم‌ساز چه تاثیری در پاسخ نهایی به دست آمده دارد و در چه شرایطی به حالت پایه ترجیح داده می‌شود؟
- در مورد  $lasso\ regression$  و  $elastic\ net$  تحقیق کنید. کاربرد هر یک چیست و مزیت هر کدام نسبت به  $ridge\ regression$  چیست؟

**بخش ۳.** فرم دیگری از مسئله *regularized least squares* که به *weighted least squares* شناخته شده است. کاربرد این روش در مسائلی است که می‌خواهیم میزان توجه متفاوتی به نقاط دیتاست‌مان داشته باشیم و نمی‌خواهیم همه نقاط به یک اندازه جواب نهایی را تحت تاثیر قرار دهند. در این مسئله تابع هدف به شکل زیر تعریف می‌شود:

$$\min_x \frac{1}{n} \sum_{i=1}^n w_i (A_i^T x - b_i)^2 \quad (w, b \in \mathbb{R}^n) \quad (۳)$$

که در آن  $A_i^T$  سطر  $i$ ام ماتریس  $A$  است. در این صورت اگر  $W$  ماتریسی قطری باشد که مقادیر قطر آن مقادیر  $w$  باشد و بخواهیم به فرم ماتریسی بازنویسی کنیم مسئله به شکل زیر می‌شود.

$$\min_x \frac{1}{n} (Ax - b)^T W (Ax - b) \quad (۴)$$

پس از محاسبه فرم بسته بردار  $x$ ، این روش را با دیتاست *california\_housing.csv* پیاده‌سازی کنید. برای بردار  $w$  سه حالت را در نظر بگیرید:

- الف) حالتی که مقادیر آن از توزیع یکنواخت (در بازه ۰.۵ تا ۳) نمونه گرفته شود،
  - ب) حالتی که مقادیر آن از توزیع چندجمله‌ای با پارامترهای  $\frac{1}{n}$  که  $n$  تعداد نمونه‌ها است نمونه گرفته شود،
  - پ) حالتی که مقادیر آن از توزیع دریشله با پارامترهای واحد (به تعداد نمونه‌ها) نمونه گرفته شود.
- توجه شود که برای بهبود نتایج هر مورد را باید در ۱۰۰ تکرار انجام دهید و بردار پاسخ را میانگین پاسخ‌های مراتب مختلف در نظر بگیرید. برای انجام نمونه‌گیری از توابع *np.random.uniform* و *np.random.multinomial* و *np.random.dirichlet* در کتابخانه *NumPy* استفاده کنید. در نهایت پاسخ نهایی و عملکرد را بین حالات مختلف مقایسه کنید.