



به نام خدا

درس یادگیری ماشین

تکلیف برنامه نویسی: رگرسیون

داده‌ها در فایل `data.npz` قرار دارند (پیوست شده است). این داده بر مبنای رابطه زیر تولید شده اند:

$$y = 4x_2^2x_1 + 2x_2^2 + 3x_1 + 1$$

بنابراین نمونه‌ها به صورت  $[x_1, x_2]$  و خروجی متناظرشان  $y$  است. در فایل `data.npz` شش آرایه یک بعدی،  $x_{2test}$ ،  $x_{1test}$ ،  $y_{test}$ ،  $x_2$ ،  $x_1$ ،  $y$  قرار دارند. یک مثال از دسترسی به آرایه  $y$  فایل چنین است:

```
a = np.load('data.npz')
```

```
print(a['y'])
```

آرایه  $y_{test}$  خروجی مطلوب روی داده های تست ( $x_{1test}, x_{2test}$ ) است و شما خروجی های خود را با این مقادیر مقایسه نمایید. در ادامه می خواهیم رگرسیون در سه مورد بررسی کنیم :

- رگرسیون خطی
- رگرسیون چندجمله ای درجه 3
- رگرسیون چندجمله ای درجه 5

در هر سه حالت تابع هزینه تابع SSE است.

(a) رگرسیون بدون منظم سازی : سه مورد را بدون جمله منظم ساز پیاده سازی کرده و نتایج را روی داده های تست با هم مقایسه کنید. از فرمول بسته رگرسیون خطی / خطی تعمیم یافته برای محاسبه  $W$  استفاده کنید.

مقدار تابع خطا روی داده ای آموزش و داده های تست را برای هر یک از سه مورد گزارش نمایید.

(b) بخش a را با استفاده از Gradient Descent انجام دهید. شرط توقف الگوریتم به عهده خودتان است.

(c) رگرسیون با منظم سازی : در این بخش به تابع هزینه SSE یک جمله منظم ساز  $L_2$  با ضریب  $\lambda$  بیفزایید. با استفاده از 5 Fold Cross Validation، بهترین مقدار پارامتر  $\lambda$  را از بین اعضای مجموعه  $10^4$ ، ...، 1،  $10^{-3}$ ،

10-4 در هر مورد بیابید و سپس نتیجه را روی داده های تست بدست آورید.