نوبران دانشگاه تبران گروه علوم کامپیوتر

به نام خدا

درس هوش مصنوعی پیشرفته (پاییز 98-97) تکلیف برنامه نویسی: رگرسیون

[رگرسیون] دادهها در فایل data.npz قرار دارند (پیوست شده است). این داده بر مبنای رابطه زیر تولیده شده اند:

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

 $x2_{test}$ بعدی، x_1,x_2 شش آرایه یک بعدی، x_1,x_2 و خروجی متناظرشان y است. در فایل y معدی، y فایل پنین است: y 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 $(x1_{test},x2_{test})$ است و شما خروجیهای خود را با این مقادیر مقایسه نمایید. در ادامه می خواهیم رگرسیون در سه مورد بررسی کنیم :

- رگرسیون خطی
- رگرسیون چندجمله ای درجه ۳
- رگرسیون چندجمله ای درجه ۵

در هر سه حالت تایع هزینه تابع SSE است.

- (a) رگرسیون بدون منظم ساری : سه مورد را بدون جمله منظمساز پیاده سازی کرده و نتایج را روی دادههای تست با هم مقایسه کنید. از فرمول بسته رگرسیون خطی I خطی تعمیم یافته برای محاسبه I استفاده کنید. مقدار تابع خطا روی داده آموزش و داده های تست را برای هر یک از سه مورد گزارش نمایید.
 - b) بخش a را باستفاده از Gradient Descent انجام دهيد. شرط توقف الگوريتم به عهده خودتان است.
- ر گرسیون با منظم سازی : در این بخش به تابع هزینه SSE یک جمله منظمساز L2 با ظریب λ بیفزایید. با استفاده از Fold Cross Validation ، بهترین مقدار پارامتر λ را ازبین اعضای مجموعه λ 10، ...، 1، 3-10 در هر مورد بیابید و سپس نتیجه را روی دادههای تست بدست آورید.