به نام خدا

هوش محاسباتی (بهار ۹۹)

تکلیف شماره ۲ : Linear and Logistic Regression & Neural Networks

---

۱. فرض کنید مجموعه داده‌ای به صورت زیر در اختیار داریم. می‌خواهیم از رگرسیون خطی با تابع هزینه SSE استفاده نمائیم و ضرائب موجود در تابع

$$Y = aX^2 + bX + c$$

را به نحوی پیدا کنیم تا تابع حاصل به بهترین حالت روی داده‌ها برازش شود.

توجه نمائید که ویژگی‌ها یک بعدی هستند و با متغیر X در جدول نشان داده شده‌اند. خروجی نیز دارای مقادیر پیوسته می‌باشد و با Y نمایش داده شده است. محاسبات را تا دو رقم اعشار انجام دهید.

(**راهنمایی:** برای به دست آوردن وزن‌ها در رگرسیون خطی از Normal Equation استفاده نمائید.)

$$\boldsymbol{SSE}:\ L = \frac{1}{2m}\sum_{i=1}^{m}\left(Y^{(i)} - X^{(i)^T}\theta\right)^2$$

X	35.38	0.29	11.74	39.45	26.85	3.98	19.05	15.32	30.51
Y	2955.53	2.28	334.32	3670.48	1709.09	13.08	864.44	560.30	2202.93

۲. مدل های رگرسیون خطی زیر را برای مساله طبقه بندی دوتایی با تابع سیگموید $g(z) = \frac{1}{1+e^{-z}}$ در نظر بگیرید:

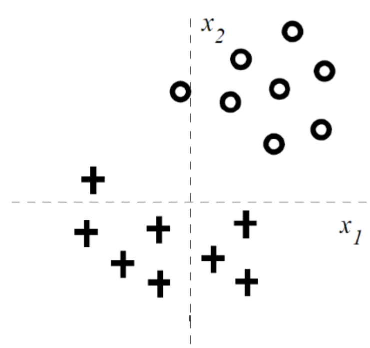
- Model 1: P (Y = 1 | X, w1, w2) = g (w1X1 + w2X2)
- Model 2: P (Y = 1 | X, w1, w2) = g (w0 + w1X1 + w2X2)

سه نمونه زیر برای آموزش این مدل موجود است:

x(1) = $[1,1]^T$ x(2) =$[1,0]^T$ x(3) = $[0,0]^T$
y(1) = 1   y(2) = −1   y(3) = 1

آیا برچسب داده سوم در مدل ۱ اهمیت دارد؟ یعنی اگر برچسب داده سوم را از ۱ به ۱- تغییر دهیم، مقادیر وزن های w = (w1, w2) که در طی آموزش یادگرفته شدند تغییر می کنند؟ آیا این تغییر برچسب در مدل ۲ اهمیت دارد؟ پاسخ های خود را به طور خلاصه شرح دهید (راهنمایی: مرز تصمیم گیری را روی یک صفحه دو بعدی در نظر بگیرید)

۳. یک روش برای حل مساله دسته بندی موجود در شکل زیر در نظر بگیرید.



هدف این است که مساله طبقه بندی باینری که در شکل بالا به تصویر کشیده شده است را با استفاده از مدل رگرسیون لاجیستیک حل کنیم.

$$P(y=1|\vec{x},\vec{w}) = g(w_0 + w_1x_1 + w_2x_2) = \frac{1}{1+\exp(-w_0 - w_1x_1 - w_2x_2)}$$

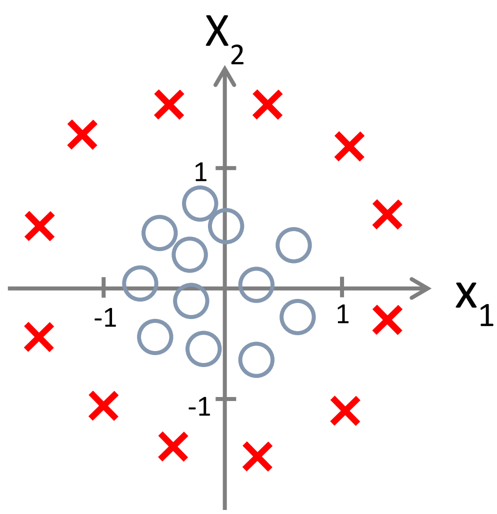
توجه داشته باشید که داده های آموزشی با خطای صفر در آموزش و با استفاده از یک جداکننده خطی در دو کلاس مختلف قرار می گیرند.
مساله آموزش این مدل رگرسیون لاجیستیک که با ضریب λ منظم شده است را در نظر بگیرید. در این مدل می خواهیم به ازای مقادیر بزرگ λ عبارت زیر را بیشینه کنیم:

$$\sum_{i=1}^{n} \log\left(P(y_i|x_i, w_0, w_1, w_2)\right) - \lambda w_j^2$$

مقدار پنالتی منتظم سازی که در این عبارت شرطی استفاده می شود برابر با $-\lambda w_j^2$ است که در آن $j = \{0,1,2\}$ می باشد. به عبارت دیگر هر بار فقط یکی از پارامترهای w_j منتظم می شوند. با توجه به داده های آموزشی تصویر شده در شکل بالا، خطای آموزش با منتظم سازی هر پارامتر w_j چه تغییری می کند؟ یعنی مشخص کنید که به ازای هر w_j برای مقادیر بزرگ λ خطای آموزش زیاد می شود یا بدون تغییر باقی می ماند. دلیل خود را به طور مختصر شرح دهید.

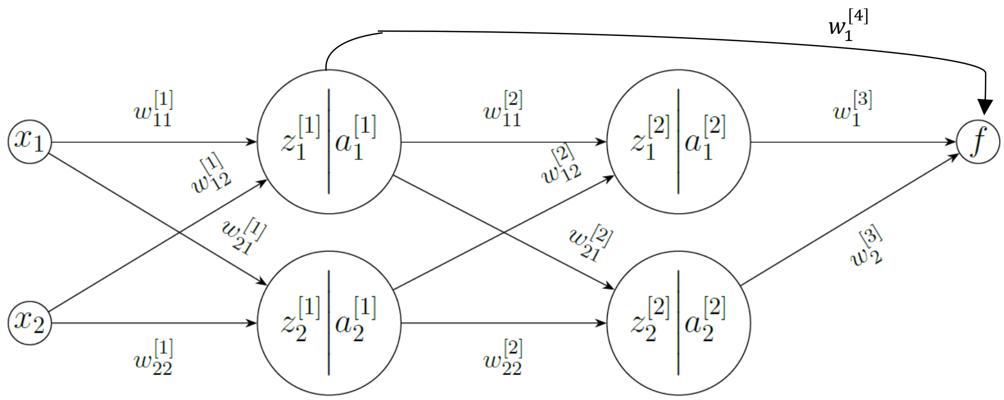
الف) منتظم سازی روی w_2.
ب) منتظم سازی روی w_1.
ج) منتظم سازی روی w_0.

٤. فرض کنید می خواهیم شبکه عصبی طراحی کنیم که داده های دو بعدی (X = [x1, x2]) به دو کلاس دسته بندی کند: کلاس دایره و کلاس ضربدر. یک مجموعه داده های آموزشی داریم که به صورت زیر است:



شبکه ای طراحی کنید که این مساله طبقه بندی را حل کند. تعداد نودها و معماری شبکه را تشریح کنید و دلایل انتخاب خود را توضیح دهید. مرز تصمیم گیری که شبکه طراحی شده توسط شما قادر به جداسازی داده ها است به چه شکل است؟
موارد خواسته شده را برای دو حالت زیر پاسخ دهید.
الف) بدون استفاده از ویژگی های quadratic
ب) با فرض وجود ویژگی های quadratic.

۵. شبکه سه لایه زیر را در نظر بگیرید:



$$Z^{[1]} = \begin{bmatrix} z_1^{[1]} \\ z_2^{[1]} \end{bmatrix} = \begin{bmatrix} w_{11}^{[1]} & w_{12}^{[1]} \\ w_{21}^{[1]} & w_{22}^{[1]} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad , \quad A^{[1]} = \begin{bmatrix} a_1^{[1]} \\ a_2^{[1]} \end{bmatrix} = \begin{bmatrix} \sigma(z_1^{[1]}) \\ \sigma(z_2^{[1]}) \end{bmatrix}$$

$$Z^{[2]} = \begin{bmatrix} z_1^{[2]} \\ z_2^{[2]} \end{bmatrix} = \begin{bmatrix} w_{11}^{[2]} & w_{12}^{[2]} \\ w_{21}^{[2]} & w_{22}^{[2]} \end{bmatrix} \begin{bmatrix} a_1^{[1]} \\ a_2^{[1]} \end{bmatrix} \quad , \quad A^{[2]} = \begin{bmatrix} a_1^{[2]} \\ a_2^{[2]} \end{bmatrix} = \begin{bmatrix} \sigma(z_1^{[2]}) \\ \sigma(z_2^{[2]}) \end{bmatrix}$$

$$f = w_1^{[3]} a_1^{[2]} + w_2^{[3]} a_2^{[2]} + w_1^{[4]} a_1^{[1]}$$

توجه شود که تابع هزینه MSE است و به صورت زیر تعریف شده است:

$$J = \frac{1}{2m} \sum_{i=1}^{m} (f^i - y^i)^2$$

مشتقات زیر را محاسبه نمایید:

a. $\delta_1 = \frac{\partial f(x)}{\partial z_1^{[2]}}$

b. $\delta_2 = \frac{\partial f(x)}{\partial Z^{[2]}}$

c. $\delta_3 = \frac{\partial f(x)}{\partial z_1^{[1]}}$

d. $\delta_4 = \frac{\partial f(x)}{\partial w_{11}^{[1]}}$

موفق باشید 😊