



به نام خدا
هوش محاسباتی (بهار ۱۴۰۱)

تکلیف شماره ۱ : Linear and Logistic Regression / Neural Networks

موعد تحویل: ۱۷ فروردین ۱۴۰۱

۱- فرض کنید برای طبقه‌بندی مجموعه داده‌های دو بعدی $X_i \in \mathbb{R}^2$ به دو کلاس صفر و یک از Logistic Regression استفاده می‌کنیم. همچنین، تابع رگرسیون، تابعی به فرم $h(z) = s(w^T z + \alpha)$ بوده که s تابع لاجستیک است. متأسفانه این روش، به خوبی داده‌ها را از یکدیگر جدا نمی‌کند. برای حل این مشکل، به هر داده، ویژگی $\|X_i\|^2$ را اضافه می‌کنیم. پس از اضافه کردن این ویژگی، مجدداً Logistic Regression را روی داده‌ها اجرا می‌کنیم. در این حالت مرز تصمیم‌گیری به چه شکل می‌شود؟ دلیل خود را شرح دهید. (امکان انتخاب چند گزینه وجود دارد)

الف) یک خط

ب) یک دایره

پ) یک بیضی

ت) یک منحنی به شکل S

۲- با توجه به جدول زیر، آیا می‌توان تابع F را با یک گره شبکه عصبی مدل کرد؟ اگر بله، وزن‌های آن را مشخص کنید، اگر خیر، استدلال خود را به طور مختصر شرح دهید.

A	B	F(A,B)
1	1	0
0	0	0
1	0	1
0	1	0

۳- مجموعه داده زیر را در نظر بگیرید و به سوالات داده شده پاسخ دهید.

Item	X1	X2	Class
A	1	2	Yes=1
B	2	1	Yes=1
C	1	1	No=0
D	1	0	No=0

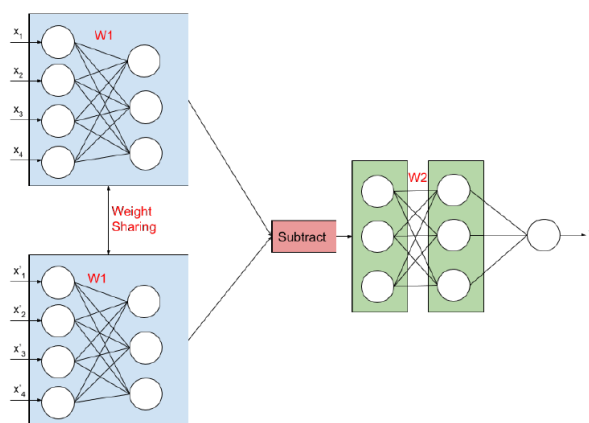
الف) آیا این مجموعه داده به صورت خطی قابل طبقه‌بندی هستند؟ دلیل خود را شرح دهید.

ب) این مجموعه داده را با یک شبکه پرسپترون تک لایه آموزش می‌دهیم. برای این کار مقدار بایاس $x_0 = -1$ را به تمام داده‌ها اضافه می‌کنیم. مقدار وزن‌ها را در مرحله اول $w = (0, -1, 1)$ و نرخ یادگیری را 0.1 در نظر بگیرید. با در نظر گرفتن نقطه A مقدار وزن‌ها چه تغییری می‌کنند؟ اگر نقطه B را در نظر بگیریم مقدار وزن‌ها چه تغییری می‌کنند؟ (برای هر نقطه محاسبه یک iteration کافی است)

پ) فرض کنید به جای استفاده از پرسپترون، از یک رگرسیون خطی برای طبقه‌بندی استفاده کنیم. در این حالت با در نظر گرفتن نقاط A و B وزن‌ها چه تغییری می‌کنند؟ نرخ یادگیری را 0.1 در نظر بگیرید. ؟ (برای هر نقطه محاسبه یک iteration کافی است)

۴- شبکه عصبی زیر را در نظر بگیرید. این شبکه، شبکه عصبی Siamese نام دارد که شامل یک شبکه دو قلو است که ورودی‌های مجزایی دارند اما وزن‌ها در این دو شبکه مشترک هستند. خروجی این شبکه دو قلو در یک یا چند لایه به هم متصل می‌شوند.

فرض کنید یک شبکه عصبی Siamese به صورت زیر داریم:



$$z_1 = W_1 x^{(i)} + b_1$$

$$a_1 = \text{ReLU}(z_1)$$

$$z_2 = W_1 x'^{(i)} + b_1$$

$$a_2 = \text{ReLU}(z_2)$$

$$a = a_1 - a_2$$

$$z_3 = W_2 a + b_2$$

$$\hat{y}^{(i)} = \sigma(z_3)$$

$$L^{(i)} = y^{(i)} \log(\hat{y}^{(i)}) + (1 - y^{(i)}) \log(1 - \hat{y}^{(i)})$$

$$J = -\frac{1}{m} \sum_{i=1}^m L^{(i)}$$

توجه کنید که $x^{(i)}$ و $x'^{(i)}$ ورودی‌های شبکه دوقلو هستند که ابعاد هر کدام از آنها $1 \times D_x$ است. همچنین خروجی مربوط به جفت ورودی نام بوده و یک عدد می‌باشد. در مجموعه داده مورد نظر، m داده وجود دارد. تعداد گره‌ها در لایه پنهان اول، D_{a1} بوده بنابراین ابعاد z_1 و z_2 ، $1 \times D_{a1}$ است. توجه کنید که در دو لایه اول وزن‌ها مشترک هستند. با توجه به اطلاعات داده شده، به سوالات زیر پاسخ دهید:

الف) ابعاد پارامترهای W_1 ، b_1 ، W_2 و b_2 به چه صورت است؟ اگر از تکنیک vectorization برای مجموعه داده استفاده کنیم، ابعاد X و Y به چه شکل می‌شود؟

ب) مقدار مشتقات زیر را حساب کنید:

$$\frac{\partial J}{\partial z_3} \bullet$$

$$\begin{array}{l} \frac{\partial a}{\partial z_2} \bullet \\ \frac{\partial J}{\partial w_1} \bullet \end{array}$$

پ) فرمول‌های به روزرسانی پارامترهای w_1 ، b_1 ، w_2 و b_2 با روش gradient descent و با نرخ یادگیری α را بنویسید. توجه کنید که مشتقات باید محاسبه شوند. می‌توانید از مقادیر مشتق بدست آمده از قسمت (ب) استفاده نمایید.

موفق باشید