دانشگاه صنعتی اصفهان

دانشکده مهندسی برق و کامپیوتر

درس مبانی هوش محاسباتی

پاسخنامه تکلیف تئوری اول

سوال ۱

- نتیجه(خروجی)
  - رگرسیون خطی: میانگین شرطی پاسخ بین $-inf$ و $+inf$ است.
  - رگرسیون لجستیک: میانگین شرطی پاسخ بین ۰ و ۱ است
- ارتباط بین متغیرها
  - رگرسیون خطی: رابطه خطی بین متغیر مستقل و وابسته
  - رگرسیون لجستیک: رابطه خطی بین شانس متغیر مستقل و لاگ متغیر وابسته
- خطا
  - رگرسیون خطی: خطای تصادفی طبیعی
  - رگرسیون لجستیک - خطای عادی تصادفی ندارد بلکه خطای دوجمله‌ای دارد ($(P \times (1-P))$)
- روش‌های تخمین [۱] (برآورد)
  - رگرسیون خطی: روش حداقل مربع معمولی (OLS)
  - رگرسیون لجستیک - روش تخمین حداکثر درستنمایی (MLe)

سؤال ۲

در رگرسیون لجستیک از تابع sigmoid استفاده می‌کنیم و یک تبدیل غیر خطی برای بدست آوردن احتمالات انجام می‌دهیم. مربع کردن این تبدیل غیر خطی منجر به عدم تحدب با حداقل‌های محلی [۲] می شود. یافتن حداقل سراسری [۳] در چنین مواردی با استفاده از نزول گرادیان امکان پذیر نیست. به همین دلیل، MSE برای رگرسیون لجستیک مناسب نیست. آنتروپی متقاطع یا از دست دادن لاگ به عنوان تابع هزینه برای رگرسیون لجستیک استفاده می‌شود. در تابع هزینه برای رگرسیون لجستیک، پیش‌بینی‌های اشتباه قطعی به شدت جریمه می‌شوند. پیش بینی های درست با ضریبا اطمینان کمتر پاداش می‌گیرند. با بهینه سازی این تابع هزینه، همگرایی حاصل می‌شود.

[1] Estimation Methods
[2] Local minimums
[3] global minimum

سؤال ۳

$$
\begin{gathered}
p(X)=\frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}} \\
1-p(X)=1-\frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}=\frac{1}{1+e^{\beta_0+\beta_1 X}} \\
\frac{P(X)}{1-P(X)}=\frac{\frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}}{\frac{1}{1+e^{\beta_0+\beta_1 X}}}=\frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}} \times \frac{1+e^{\beta_0+\beta_1 X}}{1}=e^{\beta_0+\beta_1 X} \\
\log e^{\beta_0+\beta_1 X}=\ln e^{\beta_0+\beta_1 X}=\beta_0+\beta_1 X
\end{gathered}
$$

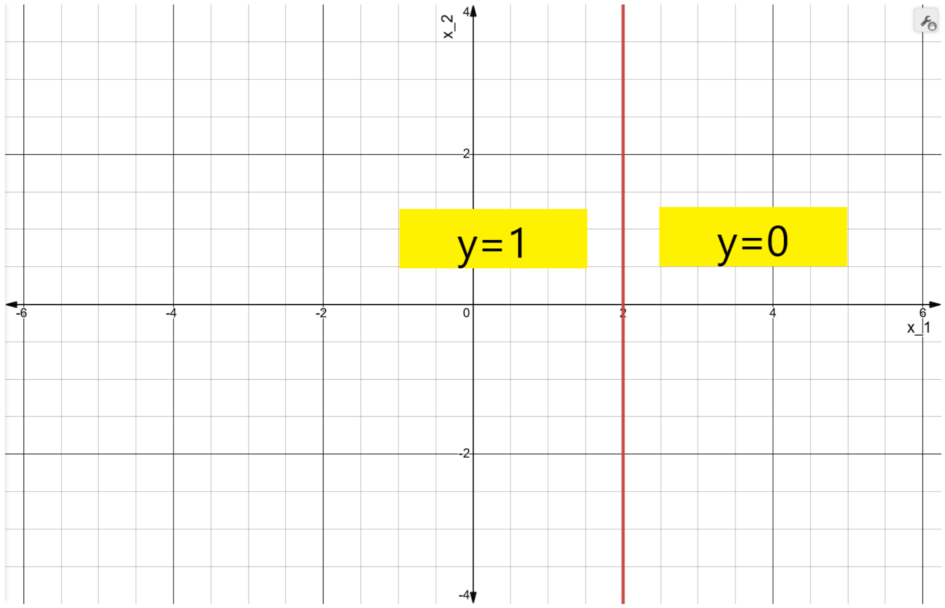
سؤال ۴

معروف ترین روش مقابله با طبقه بندی چند طبقه با استفاده از رگرسیون لجستیک، استفاده از رویکرد یک در مقابل همه [1] است. تحت این رویکرد، تعدادی مدل آموزش داده می شود که برابر با تعداد کلاس ها است. مدل ها به روش خاصی کار می‌کنند. برای مثال، مدل اول، نقطه داده را بسته به اینکه متعلق به کلاس ۱ باشد یا کلاس دیگری، طبقه بندی می‌کند. مدل دوم نقطه داده را به کلاس ۲ یا کلاس دیگری طبقه بندی می‌کند. به این ترتیب، هر نقطه داده را می‌توان در تمام کلاس‌ها بررسی کرد.

[1] One vs. all

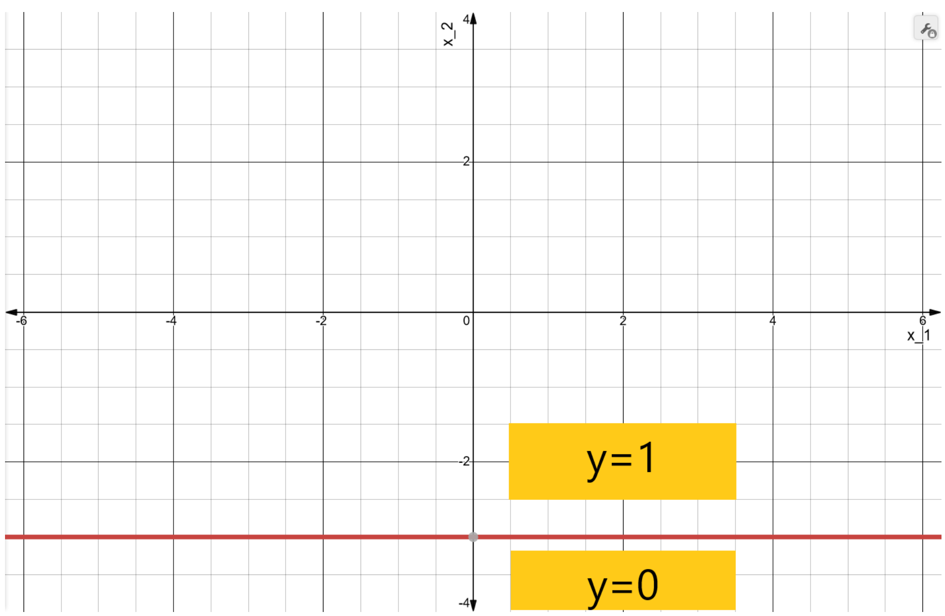
سؤال ۵

۱. $-3x_1 + 6 = 0 \Rightarrow x_1 = 2$



شکل ۱: مرز تصمیم مورد اول

۲. $6x_2 + 18 = 0 \Rightarrow x_2 = -3$



شکل ۲: مرز تصمیم مورد دوم

سؤال ۶

برای هر رنگ به طور جداگانه داریم:

۱. رنگ سیاه: $\beta_0 = 0, \beta_1 = 1$

۲. رنگ سبز: $\beta_0 = 0, \beta_1 = -1$

سؤال ۷

الف)

$$W_1 \in R^{D_{a_1} \times D_x}, W_2 \in R^{1 \times D_{a_1}}$$

$$b_1 \in R^{D_{a_1} \times 1}, b_2 \in R^{1 \times 1}$$

با استفاده از تکنیک برداری‌سازی داریم:

$$X \in R^{2D_x \times m}, Y \in R^{1 \times m}$$

ب)

- $$\begin{aligned}\frac{\partial J}{\partial z_3} &= \frac{\partial \left(-\frac{1}{m}\sum_{i=1}^{m}\left(y^{(i)}\log\left(\sigma\left(z_3\right)\right) + \left(1-y^{(i)}\right)\left(\log\left(1-\sigma\left(z_3\right)\right)\right)\right)\right)}{\partial z_3} \\ &= \frac{\partial\left(\frac{1}{m}\sum_{i=1}^{m}\left(\sigma\left(z_3\right) - y^{(i)}\right)\right)}{\partial z_3} = \frac{1}{m}\sum_{i=1}^{m}\left(\hat{y}^{(i)} - y^{(i)}\right)\end{aligned}$$

- $$\begin{aligned}\frac{\partial J}{\partial z_2} &= \frac{\partial\left(a_1 - a_2\right)}{\partial z_2} = \frac{\partial\left(\tanh z_1 - \tanh z_2\right)}{\partial z_2} = \frac{\partial\left(\tanh z_1\right)}{\partial z_2} - \frac{\partial\left(\tanh z_2\right)}{\partial z_2} = -\frac{\partial\left(\tanh z_2\right)}{\partial z_2} \\ &= -\frac{\partial\left(\frac{e^{z_2}-e^{-z_2}}{e^{z_2}+e^{-z_2}}\right)}{\partial z_2} = -\frac{\left(e^{z_2}+e^{-z_2}\right)\left(e^{z_2}+e^{-z_2}\right) - (c)\left(e^{z_2}-e^{-z_2}\right)}{\left(e^{z_2}+e^{-z_2}\right)^2} \\ &= -\left(1 - \left(\frac{\left(e^{z_2}-e^{-z_2}\right)}{\left(e^{z_2}+e^{-z_2}\right)}\right)^2\right) = -\left(1 - \left(\tanh z_2\right)^2\right) = \left(\tanh z_2\right)^2 - 1\end{aligned}$$

- $$\begin{aligned}\frac{\partial J}{\partial W_1} &= \frac{\partial J}{\partial z_3} \times \frac{\partial z_3}{\partial a} \times \frac{\partial a}{\partial W_1} \\ &= \frac{\partial J}{\partial z_3} \times \frac{\partial z_3}{\partial a} \times \left(\left(\frac{\partial a}{\partial a_1} \times \frac{\partial a_1}{\partial z_1} \times \frac{\partial z_1}{\partial W_1}\right) + \left(\frac{\partial a}{\partial a_2} \times \frac{\partial a_2}{\partial z_2} \times \frac{\partial z_2}{\partial W_1}\right)\right) \\ &= \left(\frac{1}{m}\sum_{i=1}^{m}\left(\hat{y}^{(i)} - y^{(i)}\right)\right) \times W_2 \times \left(\left(\left(\tanh z_1\right)^2 - 1\right) \times x^{(i)^T} - \left(\left(\tanh z_2\right)^2 - 1\right) \times x'^{(i)^T}\right)\end{aligned}$$

- $$\frac{\partial J}{\partial W_2} = \frac{\partial J}{\partial z_3} \times \frac{\partial z_3}{\partial W_2} = \left(\frac{1}{m}\sum_{i=1}^{m}\left(\hat{y}^{(i)} - y^{(i)}\right)\right) \times a^T$$

•
$$\begin{aligned}\frac{\partial J}{\partial b_1} &= \left(\frac{1}{m}\sum_{i=1}^{m}\left(\hat{y}^{(i)} - y^{(i)}\right)\right) \times W_2 \times \left((\tanh z_1)^2 - 1 - \left((\tanh z_2)^2 - 1\right)\right)\\ &= \left(\frac{1}{m}\sum_{i=1}^{m}\left(\hat{y}^{(i)} - y^{(i)}\right)\right) \times W_2 \times \left((\tanh z_1)^2 - (\tanh z_2)^2\right)\end{aligned}$$

•
$$\frac{\partial J}{\partial b_2} = \frac{\partial J}{\partial z_3} \times \frac{\partial z_3}{\partial b_2} = \frac{1}{m}\sum_{i=1}^{m}\left(\hat{y}^{(i)} - y^{(i)}\right)$$

ج) رابطه‌های بروزرسانی

$$W_i(t+1) = W_i(t) - \alpha\frac{\partial J}{\partial W_i}$$
$$b_i(t+1) = b_i(t) - \alpha\frac{\partial J}{\partial b_i}$$

د) این نوع از شبکه‌ها برای مقایسه میان دو چیز استفاده می‌شود، به عنوان مثال، هنگامی که بخواهیم شباهت میان دو شخص را بررسی کنیم، از یک شبکه واحد استفاده می‌کنیم. فرض کنید عکس اول برای شخص اول، عکس دوم عکس دیگری از همان شخص باشد… بنابراین برچسب ما ۱ است و این نشان‌دهنده این است که ما می‌خواهیم شبکه به این نتیجه برسد که هر دو این عکس‌ها برای یک نفر است و شباهت‌ها را پیدا کند. حال پیش بینی هرچه به یک نزدیک تر باشد پس ما به هدفمان یعنی تشخیص مشابهت دو عکس یا دو چیز دیگر نزدیک شده‌ایم: با توجه به توضیحات داده شده و مقادیر داده شده نتیجه می‌گیریم:

$$x_1' = x_1, x_2' = x_2$$

در ادامه با استفاده از تکنیک برداری‌سازی داریم:

$$X = \begin{bmatrix}1 & 2\\1 & 2\\1 & 2\\1 & 2\end{bmatrix}, Y = \begin{bmatrix}1 & 1\end{bmatrix}$$

در ادامه با استفاده از روابط محاسبات را انجام می‌دهیم:

$$z_1 = \begin{bmatrix}1 & 1\end{bmatrix} \times \begin{bmatrix}1 & 2\\1 & 2\end{bmatrix} = \begin{bmatrix}2\\4\end{bmatrix}, a_1 = \tanh z_1 = \begin{bmatrix}\tanh 2\\ \tanh 4\end{bmatrix}$$
$$z_2 = \begin{bmatrix}1 & 1\end{bmatrix} \times \begin{bmatrix}1 & 2\\1 & 2\end{bmatrix} = \begin{bmatrix}2\\4\end{bmatrix}, a_2 = \tanh z_2 = \begin{bmatrix}\tanh 2\\ \tanh 4\end{bmatrix}$$
$$a = a_1 - a_2 = \begin{bmatrix}\tanh 2\\ \tanh 4\end{bmatrix} - \begin{bmatrix}\tanh 2\\ \tanh 4\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix}, z_3 = \begin{bmatrix}1 & 1\end{bmatrix} \times \begin{bmatrix}0\\0\end{bmatrix} + \begin{bmatrix}0\end{bmatrix} = \begin{bmatrix}0\end{bmatrix}, \hat{y} = \sigma(z_3) = \begin{bmatrix}0.5\end{bmatrix}$$
$$L = 1 \times \log(0.5) + (1-1) \times \log(1-0.5) = log(0.5)$$
$$W_1 = \begin{bmatrix}1 & 1\end{bmatrix} - 1 \times (0.5-1) \times \begin{bmatrix}1\end{bmatrix} \times \begin{bmatrix}0\end{bmatrix} = \begin{bmatrix}1 & 1\end{bmatrix}, W_2 = \begin{bmatrix}1\end{bmatrix} - 1 \times (0.5-1) \times \begin{bmatrix}0\end{bmatrix} = \begin{bmatrix}1\end{bmatrix}$$
$$b_1 = \begin{bmatrix}1 & 1\end{bmatrix} - 1 \times (0.5-1) \times \begin{bmatrix}1\end{bmatrix} \times \begin{bmatrix}0\end{bmatrix} = \begin{bmatrix}1 & 1\end{bmatrix}, b_2 = \begin{bmatrix}0\end{bmatrix} - 1 \times (0.5-1) = \begin{bmatrix}-0.5\end{bmatrix}$$