مهسا امینی ۹۸۱۷۸۲۳

سوال یک

نتیجه (خروجی):

در مسائلی که میخواهیم دسته بندی انجام دهیم از logistic regression استفاده میکنیم و زمانی که میخواهیم که یک مقدار پیوسته مانند وزن را پیش بینی کنیم از linear regression استفاده میکنیم. در linear regression بهترین خطی که روی داده ها fit میشود را پیدا میکنیم ولی در logistic یک S-curve پیدا میکنیم که میتوانیم داده هارا با آن دسته بندی کنیم. پس در نهایت خروجی linear یک عدد پیوسته خواهد بود و خروجی logistic به صورت صفر و یک یا yes ، no است.

ارتباط بین متغیرها:

در linear regression ارتباط بین متغیر های مستقل و وابسته باید حتما خطی باشد ولی در logistic regression لزومی ندارد که این ارتباط خطی باشد. در linear regression ممکن است بین متغیر های مستقل collinearity وجود داشته باشد ولی در logistic regression نباید بین متغیر های مستقل collinearity وجود داشته باشد.

خطا:

برای محاسبه ی loss function در linear regression از mean square error استفاده میکنیم و برای logistic

از maximum likelihood estimation استفاده میکنیم.

روش های تخمین:

برای linear regression از least square estimation استفاده میکنیم و برای logistic از maximum likelihood استفاده میکنیم.

سوال دو

در logistic ما از تابع sigmoid استفاده میکنیم و یک تبدیل غیر خطی برای بدست آوردن احتمالات انجام میدهیم. مربع کردن این تبدیل باعث non-convexity خواهد شد و در این صورت ما نمیتوانیم بهینه های سراسری را پیدا کنیم و فقط میتوانیم بهینه ها محلی را پیدا کنیم به طور خلاصه زمانی که از MSE برای linear استفاده میکنیم تضمینی برای convex بودن وجود دارد ولی برای logistic تضمینی برای convex بودن وجود ندارد. در حالی که likelihood ، convex است.

سوال سه

$$p(X)=\frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}$$

$$\frac{p(X)}{1-p(X)}=\frac{\frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}}{1-\frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}}$$

صورت و مخرج عبارت $\frac{p(X)}{1-p(X)}$ را در $\frac{1+e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}$ ضرب میکنیم:

$$\frac{\frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}}{1-\frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}}\times\frac{\frac{1+e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}}{\frac{1+e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}}=\frac{\frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}}{\frac{1+e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}-\frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}}=\frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}-e^{\beta_0+\beta_1 X}}=e^{\beta_0+\beta_1 X}$$

پس در نهایت پس از ساده سازی به این رسیدیم:

$$\frac{p(X)}{1-p(X)}=e^{\beta_0+\beta_1 X}$$

از دو طرف ln میگیریم:

$$\ln\left(\frac{p(X)}{1-p(X)}\right)=\ln(e^{\beta_0+\beta_1 X})=\beta_0+\beta_1 X$$

سوال چهار

همان طور که میدانیم logistic یک دسته بند باینری است اما می توانیم از آن برای حل مسائل چند کلاسه نیز استفاده کنیم برای اینکار از روش one vs all استفاده میکنیم این روش به ما امکان این را میدهد که اگر بیش از دو کلاس داشتیم هم از رگرسیون logistic استفاده کنیم این روش مسئله ی چند کلاسه را به چند مسئله ی دو کلاسه تبدیل میکند یعنی اگر k کلاس داشته باشیم مسئله ی ما تبدیل به مسئله ای میشود که با k تا دسته بند باینری حل خواهد شد به عنوان مثال اگر برچسب های A، B و C را داشته باشیم به این صورت عمل میکنیم:

Model1 = A or BC

Model2 =B or AC

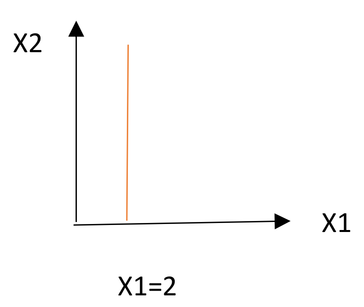
Model3 =C or AB

سوال پنج

مرز تصمیم برابر است با:

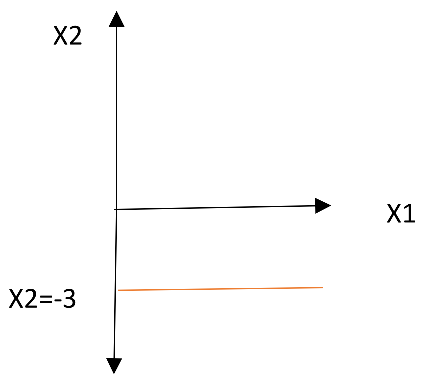
$$\theta_0+\theta_1 x_1+\theta_2 x_2=0$$

$$6-3x_1=0$$
$$3x_1=6$$
$$x_1=2$$



قسمت دو:

$$18+6x_2=0$$
$$6x_2=-18$$
$$x_2=-3$$



سوال شش

تابع سیگموید به فرم زیر است:

$$\frac{1}{1+e^{-z}}$$

زمانی که z برابر صفر است مقدار تابع برای ۰/۵ خواهد بود. حال اگر داشته باشیم :

$$z=\beta_0+\beta_1 X$$

برای مشخص کردن ضریب β_0 به شکل زیر عمل میکنیم:

بجای X صفر قرار میدهیم و در نهایت آن را در فرمول سیگموید قرار میدهیم:

$$z=\beta_0+\beta_1 * 0=\beta_0$$
$$\frac{1}{1+e^{-\beta_0}}$$

مقدار هر دو نمودار در نقطه ی صفر برابر ۰/۵ است :

$$\frac{1}{1+e^{-\beta_0}}=0.5$$
$$e^{-\beta_0}=1$$
$$\beta_0=0$$

برای مشخص کردن β_1 در نمودار سبز رنگ میتواند گفت زمانی که X منفی است مقدار آن به یک میل کرده است پس $e^{-\beta_1 X}$ باید به صفر میل کند پس β_1 باید منفی باشد . در نمودار مشکی رنگ زمانی که X مثبت است مقدار آن به یک میل میکند پس $e^{-\beta_1 X}$ باید به صفر میل کند که کافی است β_1 مثبت باشد.

به طور خلاصه مقدار β_0 برای هر دو صفر است و مقدار β_1 برای نمودار سبز منفی و برای نمودار مشکی مثبت است.

سوال هفت

قسمت الف

W_1 ماتریسی 3×4 است که 4 تعداد وروی و 3 تعداد عناصر لایه است. و به شکل زیر مشخص میشود:

$$\begin{bmatrix} W_{11} & W_{12} & W_{13} & W_{14} \\ W_{21} & W_{22} & W_{23} & W_{24} \\ W_{31} & W_{32} & W_{33} & W_{34} \end{bmatrix}$$

اگر بخواهیم W_1 را به صورت کلی در این سوال مشخص کنیم ابعاد برابر $D_{a1}\times D_x$ است.

b_1 ماتریسی است 3×1 که 3 تعداد عناصر لایه است و به شکل زیر مشخص میشود:

$$\begin{bmatrix} b_1^1 \\ b_2^1 \\ b_3^1 \end{bmatrix}$$

b_1 به صورت کلی در این سوال برابر $D_{a1}\times1$ است.

W_2 ماتریسی 1×3 است که 3 تعداد عناصر ورودی و 1 تعداد عناصر لایه است و به شکل زیر مشخص میشود:

$$\begin{bmatrix} W_{11} & W_{12} & W_{13} \end{bmatrix}$$

W_2 برابر $1\times D_{a1}$ است.

همچنین b_2 نیز 1×1 است زیرا در این لایه فقط یک پرسپترون داریم.

مشخص کردن ابعاد X وY

برای x_i یک داده است داریم:

$$z_1 = W_1 x^{(i)} + b_1$$
$$D_{a1}\times1 = (D_{a1}\times D_x)x^{(i)} + D_{a1}\times1$$

پس ابعاد $x^{(i)}$ برابر $D_x\times1$ خواهد شد و از آنجایی که m تا داده داریم ابعاد X برابر $D_x\times m$ است .

ابعاد Y هم برابر $1\times m$ خواهد بود.

قسمت ب

$$-\frac{\partial J}{\partial z_3}=\frac{\partial J}{\partial L}\times\frac{\partial L}{\partial \hat{y}}\times\frac{\partial \hat{y}}{\partial z_3}$$

$$\frac{\partial J}{\partial L}=-m$$

$$\frac{\partial L}{\partial \hat{y}}=\frac{y}{\hat{y}}-\frac{1-y}{1-\hat{y}}=\frac{y-\hat{y}}{\hat{y}(1-\hat{y})}$$

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

$$\frac{\partial \hat{y}}{\partial z_3}\to \hat{y}(1-\hat{y})$$

$$\frac{\partial J}{\partial z_3}=-m\times\frac{y-\hat{y}}{\hat{y}(1-\hat{y})}\times\hat{y}(1-\hat{y})=m(\hat{y}-y)$$

$$\frac{\partial a}{\partial z_2}=\frac{\partial a_1}{\partial z_2}-\frac{\partial a_2}{\partial z_2}=0--\frac{\partial a_2}{\partial z_2}$$

$$\frac{\partial a_2}{\partial z_2}\to$$

$$a_2=\tanh(z_2)$$

$$\tanh^{'}(z_2)=1-\tanh^2(z_2)$$

$$\frac{\partial a_2}{\partial z_2}=-(1-a_2^2)=a_2^2-1$$

$$\frac{\partial J}{\partial W_1}=\frac{\partial J}{\partial L}\times\frac{\partial L}{\partial \hat{y}}\times\frac{\partial \hat{y}}{\partial z_3}\times\frac{\partial z_3}{\partial a}\times(\frac{\partial a}{\partial a_1}\times\frac{\partial a_1}{\partial z_1}\times\frac{\partial z_1}{\partial W_1}+\frac{\partial a}{\partial a_2}\times\frac{\partial a_2}{\partial z_2}\times\frac{\partial z_2}{\partial W_1})$$

$$\frac{\partial J}{\partial L}\times\frac{\partial L}{\partial \hat{y}}\times\frac{\partial \hat{y}}{\partial z_3}=m(\hat{y}-y)$$

$$\frac{\partial z_3}{\partial a}=W_2$$

$$\frac{\partial a}{\partial a_1}\times\frac{\partial a_1}{\partial z_1}\times\frac{\partial z_1}{\partial W_1}=1\times(1-a_1{}^2)\times x$$

$$\frac{\partial a}{\partial a_2}\times\frac{\partial a_2}{\partial z_2}\times\frac{\partial z_2}{\partial W_1}=-1\times(1-a_2{}^2)\times x'$$

$$\frac{\partial J}{\partial W_1}=m(\hat{y}-y)\times W_2\times((1-a_1{}^2)\times x-(1-a_2{}^2)\times x')$$

قسمت ج

$$W_1 = W_1 - \alpha \frac{\partial J}{\partial W_1}$$

$$W_1 = W_1 - \alpha(m(\hat{y} - y) \times W_2 \times ((1 - a_1{}^2) \times x - (1 - a_2{}^2) \times x')))$$

$$b_1 = b_1 - \alpha \frac{\partial J}{\partial b_1}$$

$$\frac{\partial J}{\partial b_1} = \frac{\partial J}{\partial L} \times \frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial z_3} \times \frac{\partial z_3}{\partial a} \times (\frac{\partial a}{\partial a_1} \times \frac{\partial a_1}{\partial z_1} \times \frac{\partial z_1}{\partial b_1} + \frac{\partial a}{\partial a_2} \times \frac{\partial a_2}{\partial z_2} \times \frac{\partial z_2}{\partial b_1})$$

$$\frac{\partial J}{\partial L} \times \frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial z_3} = m(\hat{y} - y)$$

$$\frac{\partial z_3}{\partial a} = W_2$$

$$\frac{\partial a}{\partial a_1} \times \frac{\partial a_1}{\partial z_1} \times \frac{\partial z_1}{\partial b_1} = 1 \times (1 - a_1{}^2)$$

$$\frac{\partial a}{\partial a_2} \times \frac{\partial a_2}{\partial z_2} \times \frac{\partial z_2}{\partial b_1} = -1 \times (1 - a_2{}^2)$$

$$\frac{\partial J}{\partial b_1} = (\hat{y} - y) \times W_2 \times ((1 - a_1{}^2) - (1 - a_2{}^2)))$$

$$b_1 = b_1 - \alpha(m(\hat{y} - y) \times W_2 \times ((1 - a_1{}^2) - (1 - a_2{}^2))))$$

$$W_2 = W_2 - \alpha \frac{\partial J}{\partial W_2}$$

$$\frac{\partial J}{\partial W_2} = \frac{\partial J}{\partial L} \times \frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial z_3} \times \frac{\partial z_3}{\partial W_2}$$

$$\frac{\partial J}{\partial L} \times \frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial z_3} = m(\hat{y} - y)$$

$$\frac{\partial z_3}{\partial W_2} = a$$

$$W_2 = W_2 - \alpha(m(\hat{y} - y)a)$$

$$b_2 = b_2 - \alpha \frac{\partial J}{\partial b_2}$$

$$\frac{\partial J}{\partial b_2} = \frac{\partial J}{\partial L} \times \frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial z_3} \times \frac{\partial z_3}{\partial b_2}$$

$$\frac{\partial J}{\partial L} \times \frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial z_3} = m(\hat{y} - y)$$

$$\frac{\partial z_3}{\partial b_2} = 1$$

$$b_2 = b_2 - \alpha m(\hat{y} - y)$$

قسمت د

$$W_1 = W_1 - \alpha(m(\hat{y} - y) \times W_2 \times ((1 - a_1{}^2) \times x - (1 - a_2{}^2) \times x')))$$

$$b_1 = b_1 - \alpha(m(\hat{y} - y) \times W_2 \times ((1 - a_1{}^2) - (1 - a_2{}^2))))$$

$$W_2 = W_2 - \alpha(m(\hat{y} - y)a)$$

$$b_2 = b_2 - \alpha m(\hat{y} - y)$$

$$\hat{y} = \sigma(z_3) = \sigma(w_2 a + b_2)$$

$$a = a_1 - a_2 = \tanh(w_1 x + b_1) - \tanh(w_1 x' + b_1)$$

$$w_1 x + b_1 = 2$$

$$\tanh(2) = 0.96$$

$$w_1 x' + b_1 = 4$$

$$\tanh(4) = 0.99$$

$$a = -0.03$$

$$\hat{y}^1 = \sigma(-0.03) = 0.492$$

$$W_1 = (1,1) - 1((0.492 - 1) \times 1 \times ((1 - 0.96^2) \times (1,1) - (1 - 0.99^2) \times (2,2)))) = (1.0007112, 1.0007112)$$

$$b_1 = 0 - 1((0.492 - 1) \times 1 \times ((1 - 0.96^2) - (1 - 0.99^2)))) = 0.000762$$

$$W_2 = 1 - 1((0.492 - 1)a) = 0.98476$$

$$b_2 = 0 - 1(0.492 - 1) = 0.508$$