logistic regression in matrix algebra

Chowdhury Mofizur Rahman

July 2025

1 Finding classes of m examples in binary classification

Suppose we have a data set having f number of attributes $x_1, x_2, ...x_f$ and we will find out the classes of m number of examples using the z score $z = w_1x_1 + w_2x_2 + w_3x_3 + ... + w_fx_f + b$ where all the ws and b are known. We will have a matrix having m rows and f columns to represent the m examples. We will have another matrix f rows and single column to represent the ws. If we multiply these two matrix having dimensions $m \times f$ and $f \times 1$, we will get an output matrix having dimension $m \times 1$. Every row represents the summation $w_1x_1 + w_2x_2 + w_3x_3 + ... + w_fx_f$ for each example. Now we add this $m \times 1$ matrix with $m \times 1$ column matrix (each element represents b) to get $m \times 1$ matrix representing the z values of each example. Applying σ function on each element thus gives the probability of belonging to $y = 1$ class for each of the m examples.

$$\begin{bmatrix} {x_1}^1 & {x_2}^1 & {x_1}^3 .... & {x_f}^1 \\ {x_1}^2 & {x_2}^2 & {x_3}^2 .... & {x_f}^2 \\ \dots & \dots & \dots .... & \dots \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ {x_1}^m & {x_2}^m & {x_3}^m .... & {x_f}^m \end{bmatrix} \times \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \vdots \\ w_f \end{bmatrix} = \begin{bmatrix} {x_1}^1 w_1 + {x_2}^1 w_2 + {x_1}^3 w_3 + \ldots + {x_f}^1 w_f \\ {x_1}^2 w_1 + {x_2}^2 w_2 + {x_3}^2 w_3 + \ldots + {x_f}^2 w_f \\ {x_1}^3 w_1 + {x_2}^3 w_2 + {x_3}^3 w_3 + \ldots + {x_f}^3 w_f \\ \vdots \\ {x_1}^m w_1 + {x_2}^m w_2 + {x_3}^m w_3 + \ldots + {x_f}^m w_f \end{bmatrix}$$

$$\sigma\left(\begin{bmatrix} {x_1}^1 w_1 + {x_2}^1 w_2 + {x_1}^3 w_3 + \ldots + {x_f}^1 w_f \\ {x_1}^2 w_1 + {x_2}^2 w_2 + {x_3}^2 w_3 + \ldots + {x_f}^2 w_f \\ {x_1}^3 w_1 + {x_2}^3 w_2 + {x_3}^3 w_3 + \ldots + {x_f}^3 w_f \\ \vdots \\ {x_1}^m w_1 + {x_2}^m w_2 + {x_3}^m w_3 + \ldots + {x_f}^m w_f \end{bmatrix} + \begin{bmatrix} b \\ b \\ b \\ \\ b \end{bmatrix}\right) = \begin{bmatrix} \hat{y_1} \\ \hat{y_2} \\ \hat{y_3} \\ \vdots \\ \hat{y_m} \end{bmatrix}$$

Now to update weights for a batch of m training examples, we can use the following matrix algebra to updates the weights in a single matrix multiplication

$$\begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \vdots \\ w_f \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \vdots \\ w_f \end{bmatrix} - \eta \times \frac{1}{m} \left(\left(\begin{bmatrix} \hat{y_1} \\ \hat{y_2} \\ \hat{y_3} \\ \vdots \\ \hat{y_m} \end{bmatrix} - \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_m \end{bmatrix} \right)^T \times \begin{bmatrix} {x_1}^1 & {x_2}^1 & {x_3}^1 .... & {x_f}^1 \\ {x_1}^2 & {x_2}^2 & {x_3}^2 .... & {x_f}^2 \\ \ldots & \ldots & \ldots .... & \ldots \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ {x_1}^m & {x_2}^m & {x_3}^m .... & {x_f}^m \end{bmatrix} \right)^T$$

$$= \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \vdots \\ w_f \end{bmatrix} - \eta \times \frac{1}{m} \left(\left(\begin{bmatrix} \hat{y_1} - y_1 \\ \hat{y_2} - y_2 \\ \hat{y_3} - y_3 \\ \vdots \\ \hat{y_m} - y_m \end{bmatrix} \right)^T \times \begin{bmatrix} {x_1}^1 & {x_2}^1 & {x_3}^1 .... & {x_f}^1 \\ {x_1}^2 & {x_2}^2 & {x_3}^2 .... & {x_f}^2 \\ \ldots & \ldots & \ldots .... & \ldots \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ {x_1}^m & {x_2}^m & {x_3}^m .... & {x_f}^m \end{bmatrix} \right)^T$$

$$= \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \vdots \\ w_f \end{bmatrix} - \eta \times \frac{1}{m} \begin{bmatrix} (\hat{y_1} - y_1){x_1}^1 + (\hat{y_2} - y_2){x_1}^2 + (\hat{y_3} - y_3){x_1}^3 + ... + (\hat{y_m} - y_m){x_1}^m \\ (\hat{y_1} - y_1){x_2}^1 + (\hat{y_2} - y_2){x_2}^2 + (\hat{y_3} - y_3){x_2}^3 + ... + (\hat{y_m} - y_m){x_2}^m \\ (\hat{y_1} - y_1){x_3}^1 + (\hat{y_2} - y_2){x_3}^2 + (\hat{y_3} - y_3){x_3}^3 + ... + (\hat{y_m} - y_m){x_3}^m \\ \vdots \\ (\hat{y_1} - y_1){x_f}^1 + (\hat{y_2} - y_2){x_f}^2 + (\hat{y_3} - y_3){x_f}^3 + ... + (\hat{y_m} - y_m){x_f}^m \end{bmatrix}$$

2 Finding classes of m examples in k-ary classification

Suppose we have a data set having f number of attributes $x_1, x_2, ...x_f$ and we will find out the classes of m number of examples where each example belongs to one of k classes. We have k number of z score $z_1, z_2, z_3, ..., z_k$ for k different classes. Each z_i is expressed for j-th example by $z_i = w_{i1}{x_1}^j + w_{i2}{x_2}^j + w_{i3}{x_3}^j + ... + w_{if}{x_f}^j + b_i$ where all the w_{ij}s and b_i are known. We will have a matrix having m rows and f columns to represent the m examples. We will have another matrix f rows and k columns to represent the w_{ij}s. If we multiply these two matrix having dimensions $m \times f$ and $f \times k$, we will get an output matrix having dimension $m \times k$. Every element represents the summation $w_{i1}{x_1}^j + w_{i2}{x_2}^j + w_{i3}{x_3}^j + ... + w_{if}{x_f}^j$ for z_i score of each j-th example. Now we add this $m \times k$ matrix with $m \times k$ (each row represents b_i for $i = 1$ to $i = k$) to get $m \times k$ matrix where each row represents the z_i values of each i-th example. Applying function on each element of each row thus gives the probability of belonging to $y_k = 1$ class for each of the m examples.

$$\begin{bmatrix} {x_1}^1 & {x_2}^1 & {x_1}^3..... & {x_f}^1 \\ {x_1}^2 & {x_2}^2 & {x_3}^2..... & {x_f}^2 \\ \dots & \dots & \dots..... & \dots \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ {x_1}^m & {x_2}^m & {x_3}^m..... & {x_f}^m \end{bmatrix} \times \begin{bmatrix} w_{11} & w_{21} & w_{\{}31 & .... & w_{k1} \\ w_{12} & w_{22} & w_{32} & ..... & w_{k2} \\ w_{13} & w_{23} & w_{33} & ..... & w_{k3} \\ \vdots & \vdots & \vdots & .... & \vdots \\ w_{1f} & w_{2f} & w_{3f} & .... & w_{kf} \end{bmatrix}$$

$$= \begin{bmatrix} {x_1}^1 w_{11} + {x_2}^1 w_{12} + {x_3}^1 w_{13} + \ldots + {x_f}^1 w_{1f} & {x_1}^1 w_{21} + {x_2}^1 w_{22} + {x_3}^1 w_{23} + \ldots + {x_f}^1 w_{2f} & \dots & {x_1}^1 w_{k1} + {x_2}^1 w_{k2} + {x_3}^1 w_{k3} + \ldots + {x_f}^1 w_{kf} \\ {x_1}^2 w_{11} + {x_2}^2 w_{12} + {x_3}^2 w_{13} + \ldots + {x_f}^2 w_{1f} & {x_1}^2 w_{21} + {x_2}^2 w_{22} + {x_3}^2 w_{23} + \ldots + {x_f}^2 w_{2f} & \dots & {x_1}^2 w_{k1} + {x_2}^2 w_{k2} + {x_3}^2 w_{k3} + \ldots + {x_f}^2 w_{kf} \\ {x_1}^3 w_{11} + {x_2}^3 w_{12} + {x_3}^3 w_{13} + \ldots + {x_f}^3 w_{1f} & {x_1}^3 w_{21} + {x_2}^3 w_{22} + {x_3}^3 w_{23} + \ldots + {x_f}^3 w_{2f} & \dots & {x_1}^1 w_{k1} + {x_2}^1 w_{k2} + {x_3}^1 w_{k3} + \ldots + {x_f}^1 w_{kf} \\ \vdots & \vdots & \dots & \vdots \\ {x_1}^m w_{11} + {x_2}^m w_{12} + {x_3}^m w_{13} + \ldots + {x_f}^m w_{1f} & {x_1}^m w_{21} + {x_2}^m w_{22} + {x_3}^m w_{23} + \ldots + {x_f}^m w_{2f} & \dots & {x_1}^m w_{k1} + {x_2}^m w_{k2} + {x_3}^m w_{k3} + \ldots + {x_f}^m w_{kf} \end{bmatrix}$$

$$\begin{bmatrix} {x_1}^1 w_{11} + {x_2}^1 w_{12} + {x_3}^1 w_{13} + \ldots + {x_f}^1 w_{1f} & {x_1}^1 w_{21} + {x_2}^1 w_{22} + {x_3}^1 w_{23} + \ldots + {x_f}^1 w_{2f} & \dots & {x_1}^1 w_{k1} + {x_2}^1 w_{k2} + {x_3}^1 w_{k3} + \ldots + {x_f}^1 w_{kf} \\ {x_1}^2 w_{11} + {x_2}^2 w_{12} + {x_3}^2 w_{13} + \ldots + {x_f}^2 w_{1f} & {x_1}^2 w_{21} + {x_2}^2 w_{22} + {x_3}^2 w_{23} + \ldots + {x_f}^2 w_{2f} & \dots & {x_1}^2 w_{k1} + {x_2}^2 w_{k2} + {x_3}^2 w_{k3} + \ldots + {x_f}^2 w_{kf} \\ {x_1}^3 w_{11} + {x_2}^3 w_{12} + {x_3}^3 w_{13} + \ldots + {x_f}^3 w_{1f} & {x_1}^3 w_{21} + {x_2}^3 w_{22} + {x_3}^3 w_{23} + \ldots + {x_f}^3 w_{2f} & \dots & {x_1}^1 w_{k1} + {x_2}^1 w_{k2} + {x_3}^1 w_{k3} + \ldots + {x_f}^1 w_{kf} \\ \vdots & \vdots & \dots & \vdots \\ {x_1}^m w_{11} + {x_2}^m w_{12} + {x_3}^m w_{13} + \ldots + {x_f}^m w_{1f} & {x_1}^m w_{21} + {x_2}^m w_{22} + {x_3}^m w_{23} + \ldots + {x_f}^m w_{2f} & \dots & {x_1}^m w_{k1} + {x_2}^m w_{k2} + {x_3}^m w_{k3} + \ldots + {x_f}^m w_{kf} \end{bmatrix}$$

$$+ \begin{bmatrix} b_1 & b_2 & b_3 & \dots & b_k \\ b_1 & b_2 & b_3 & \dots & b_k \\ b_1 & b_2 & b_3 & \dots & b_k \\ \vdots & \vdots & \vdots & \dots & \vdots \\ b_1 & b_2 & b_3 & \dots & b_k \end{bmatrix} = \begin{bmatrix} {z_1}^1 & {z_2}^1 & {z_3}^1 & \dots & {z_k}^1 \\ {z_1}^2 & {z_2}^2 & {z_3}^2 & \dots & {z_k}^2 \\ {z_1}^3 & {z_2}^3 & {z_3}^3 & \dots & {z_k}^3 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ {z_1}^m & {z_2}^m & {z_3}^m & \dots & {z_k}^m \end{bmatrix}$$

$$\text{Softmax}\left(\begin{bmatrix} {z_1}^1 & {z_2}^1 & {z_3}^1 & \dots & {z_k}^1 \\ {z_1}^2 & {z_2}^2 & {z_3}^2 & \dots & {z_k}^2 \\ {z_1}^3 & {z_2}^3 & {z_3}^3 & \dots & {z_k}^3 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ {z_1}^m & {z_2}^m & {z_3}^m & \dots & {z_k}^m \end{bmatrix} \right) = \begin{bmatrix} \hat{y}^1 \\ \hat{y}^2 \\ \hat{y}^3 \\ \vdots \\ \hat{y}^m \end{bmatrix}$$