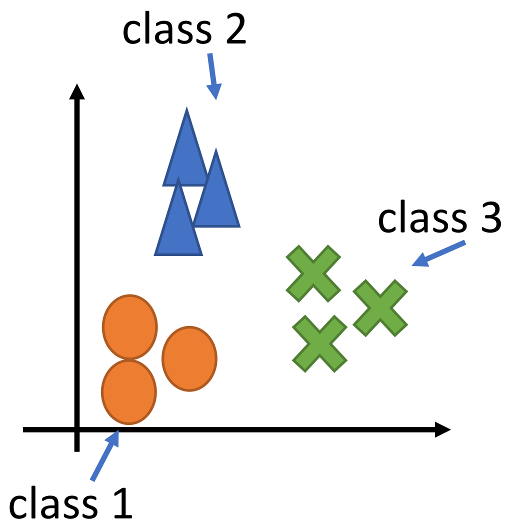
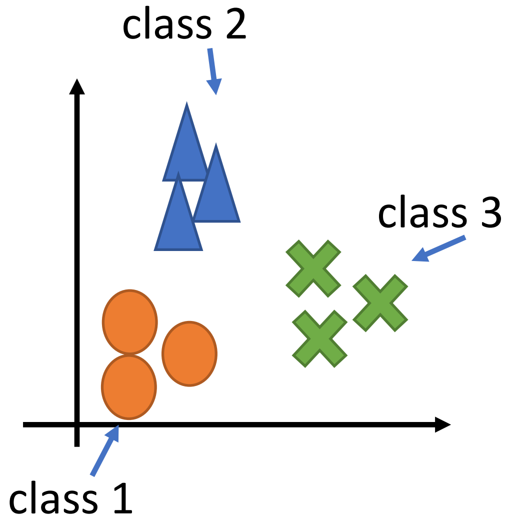
Multi-Class Classification

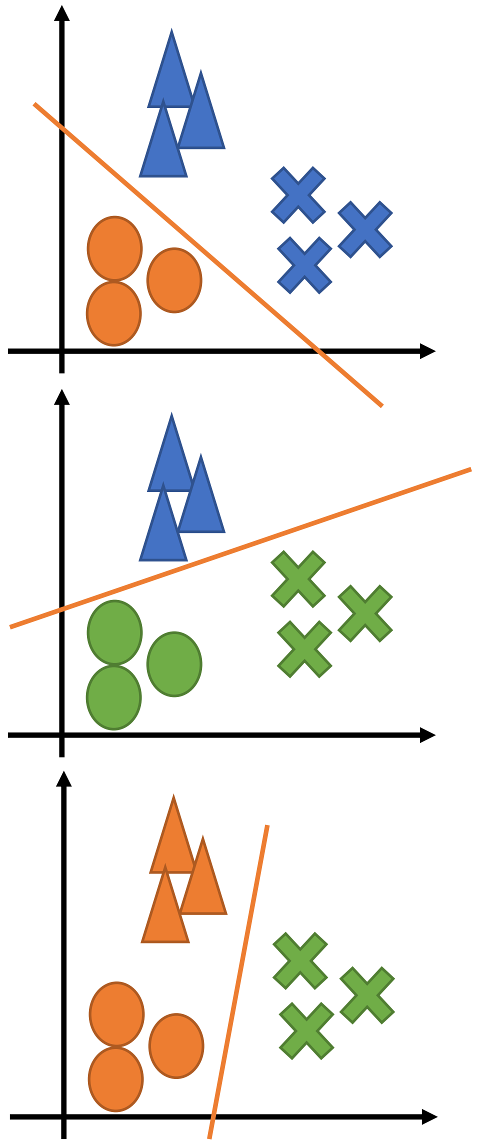
- One vs All



Multi-Class Classification

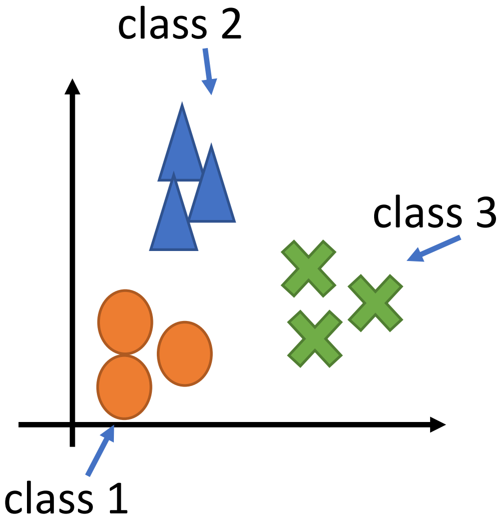
- One vs All

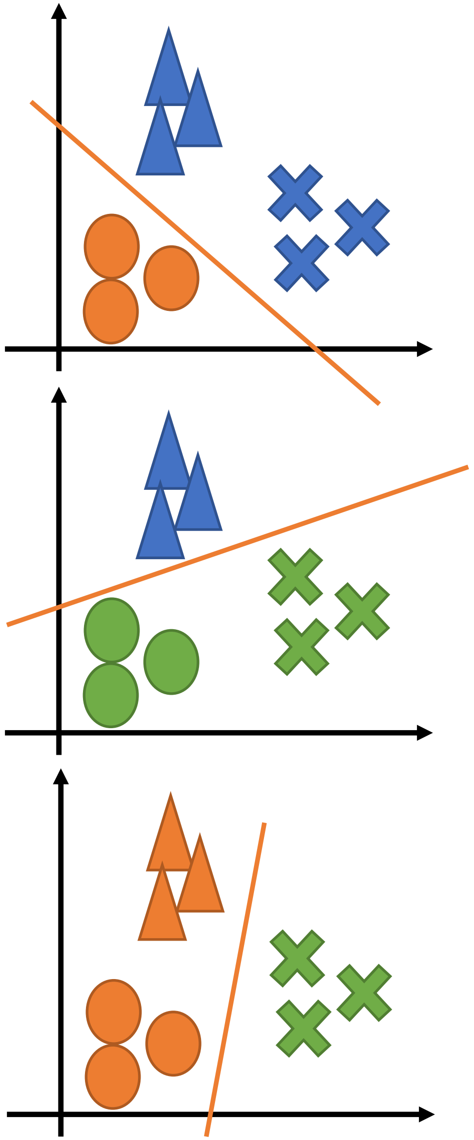




Multi-Class Classification

- One vs All





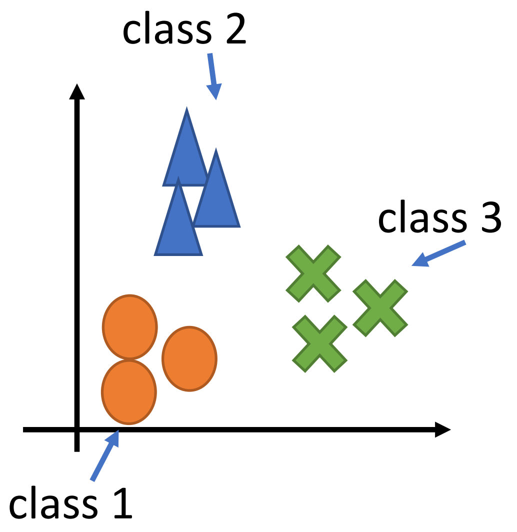
Model 1: probability in class 1

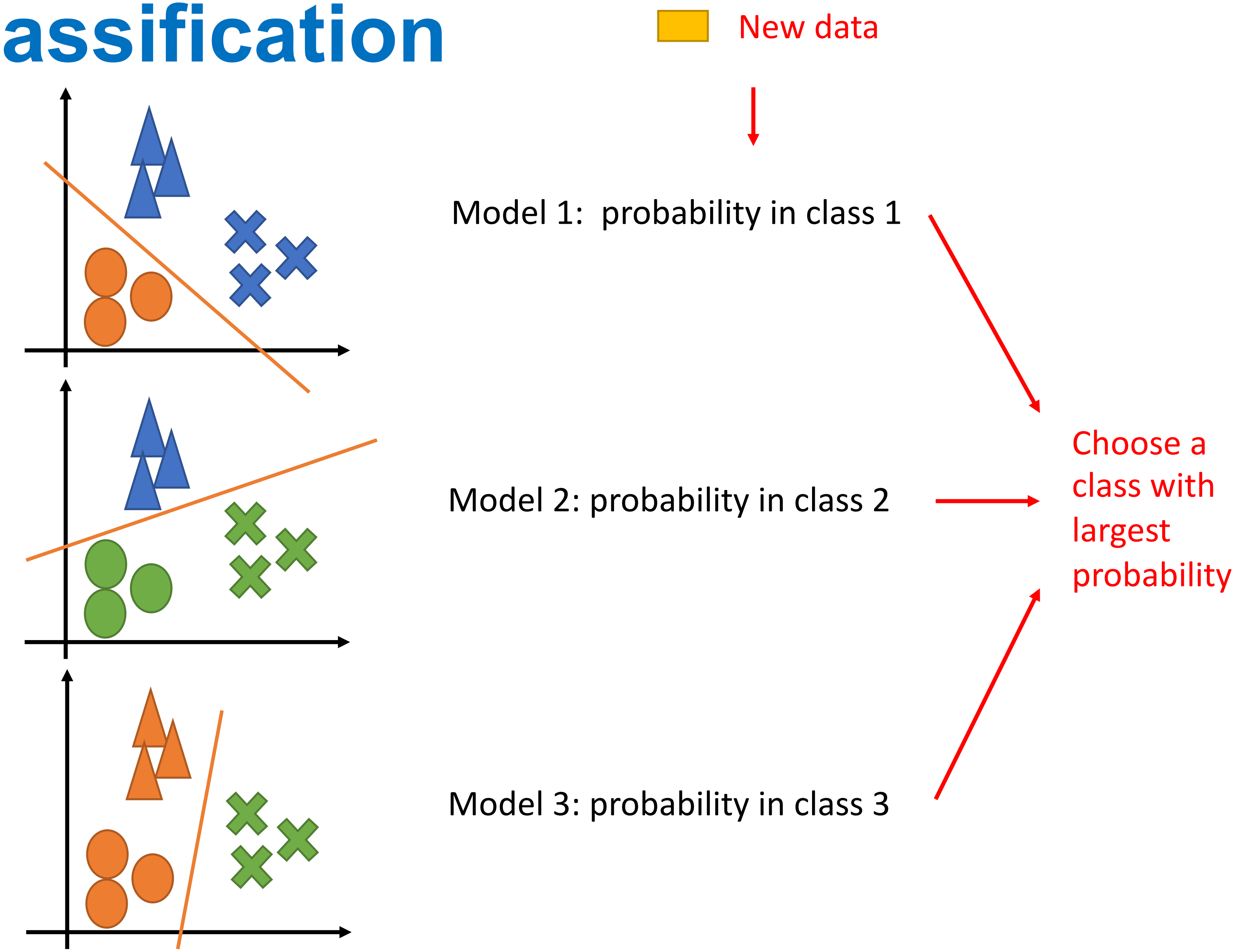
Model 2: probability in class 2

Model 3: probability in class 3

Multi-Class Classification

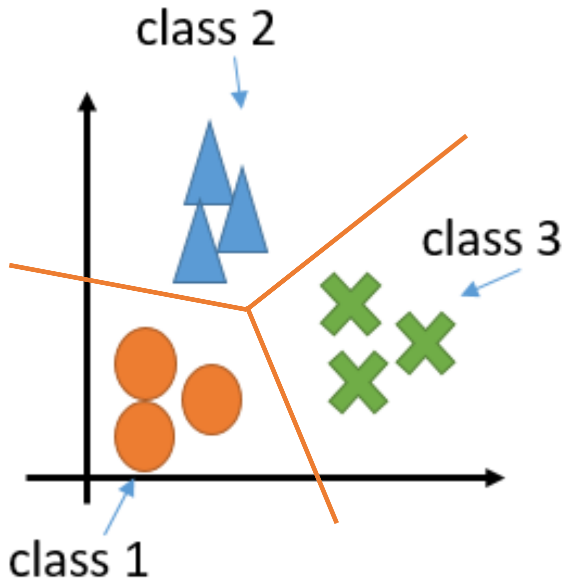
- One vs All





Multi-Class Classification

- We can also consider multi classes at the same time



C1: $w^{(1)^T}x \longrightarrow e^{w^{(1)^T}x}$

C2: $w^{(2)^T}x \longrightarrow e^{w^{(2)^T}x}$

C3: $w^{(3)^T}x \longrightarrow e^{w^{(3)^T}x}$

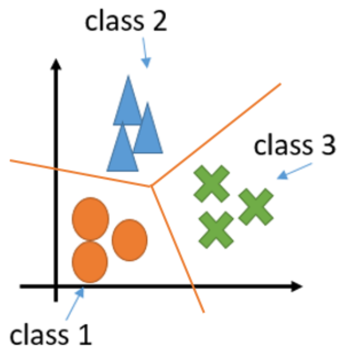
$$y_1 = \frac{e^{w^{(1)^T}x}}{e^{w^{(1)^T}x} + e^{w^{(2)^T}x} + e^{w^{(3)^T}x}}$$

$$y_2 = \frac{e^{w^{(2)^T}x}}{e^{w^{(1)^T}x} + e^{w^{(2)^T}x} + e^{w^{(3)^T}x}}$$

$$y_3 = \frac{e^{w^{(3)^T}x}}{e^{w^{(1)^T}x} + e^{w^{(2)^T}x} + e^{w^{(3)^T}x}}$$

Courtesy of Dr. Hung-yi Lee

Multi-Class Classification



- We can also consider multi classes at the same time

C1: $w^{(1)^T}x \longrightarrow e^{w^{(1)^T}x}$

C2: $w^{(2)^T}x \longrightarrow e^{w^{(2)^T}x}$

C3: $w^{(3)^T}x \longrightarrow e^{w^{(3)^T}x}$

$$y_1 = \frac{e^{w^{(1)^T}x}}{e^{w^{(1)^T}x} + e^{w^{(2)^T}x} + e^{w^{(3)^T}x}} \qquad \hat{y}_1$$

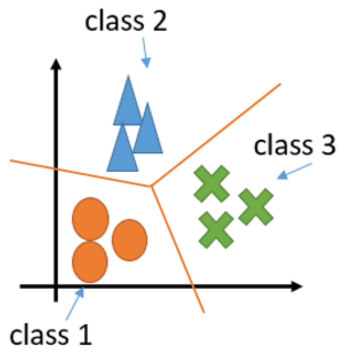
$$y_2 = \frac{e^{w^{(2)^T}x}}{e^{w^{(1)^T}x} + e^{w^{(2)^T}x} + e^{w^{(3)^T}x}} \Longleftrightarrow \hat{y}_2$$

$$y_3 = \frac{e^{w^{(3)^T}x}}{e^{w^{(1)^T}x} + e^{w^{(2)^T}x} + e^{w^{(3)^T}x}} \qquad \hat{y}_3$$

In Optimization: the loss function is $-\sum_{i=1}^{3} \hat{y}_i \, ln y_i$

Courtesy of Dr. Hung-yi Lee

Multi-Class Classification



- We can also consider multi classes at the same time

C1: $w^{(1)^T}x \longrightarrow e^{w^{(1)^T}x}$

C2: $w^{(2)^T}x \longrightarrow e^{w^{(2)^T}x}$

C3: $w^{(3)^T}x \longrightarrow e^{w^{(3)^T}x}$

$$y_1 = \frac{e^{w^{(1)^T}x}}{e^{w^{(1)^T}x} + e^{w^{(2)^T}x} + e^{w^{(3)^T}x}} \qquad \hat{y}_1$$

$$y_2 = \frac{e^{w^{(2)^T}x}}{e^{w^{(1)^T}x} + e^{w^{(2)^T}x} + e^{w^{(3)^T}x}} \iff \hat{y}_2$$

$$y_3 = \frac{e^{w^{(3)^T}x}}{e^{w^{(1)^T}x} + e^{w^{(2)^T}x} + e^{w^{(3)^T}x}} \qquad \hat{y}_3$$

If x is in C1,

the output $\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$

If x is in C2,

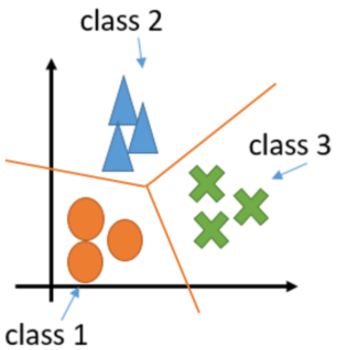
the output $\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$

If x is in C3,

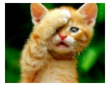
the output $\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

Courtesy of Dr. Hung-yi Lee

Multi-Class Classification



- Consider multi classes at the same time



x

$w^{(1)^T}x \rightarrow e^{w^{(1)^T}x}$

$w^{(2)^T}x \rightarrow e^{w^{(2)^T}x}$

$w^{(3)^T}x \rightarrow e^{w^{(3)^T}x}$

$$y_1 = \frac{e^{w^{(1)^T}x}}{e^{w^{(1)^T}x} + e^{w^{(2)^T}x} + e^{w^{(3)^T}x}} \qquad \hat{y}_1$$

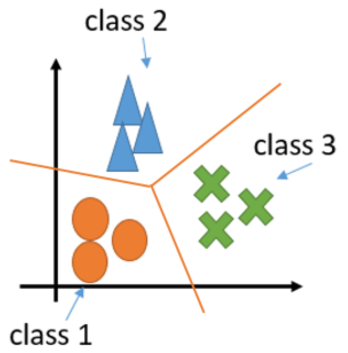
$$y_2 = \frac{e^{w^{(2)^T}x}}{e^{w^{(1)^T}x} + e^{w^{(2)^T}x} + e^{w^{(3)^T}x}} \qquad \hat{y}_2$$

$$y_3 = \frac{e^{w^{(3)^T}x}}{e^{w^{(1)^T}x} + e^{w^{(2)^T}x} + e^{w^{(3)^T}x}} \qquad \hat{y}_3$$

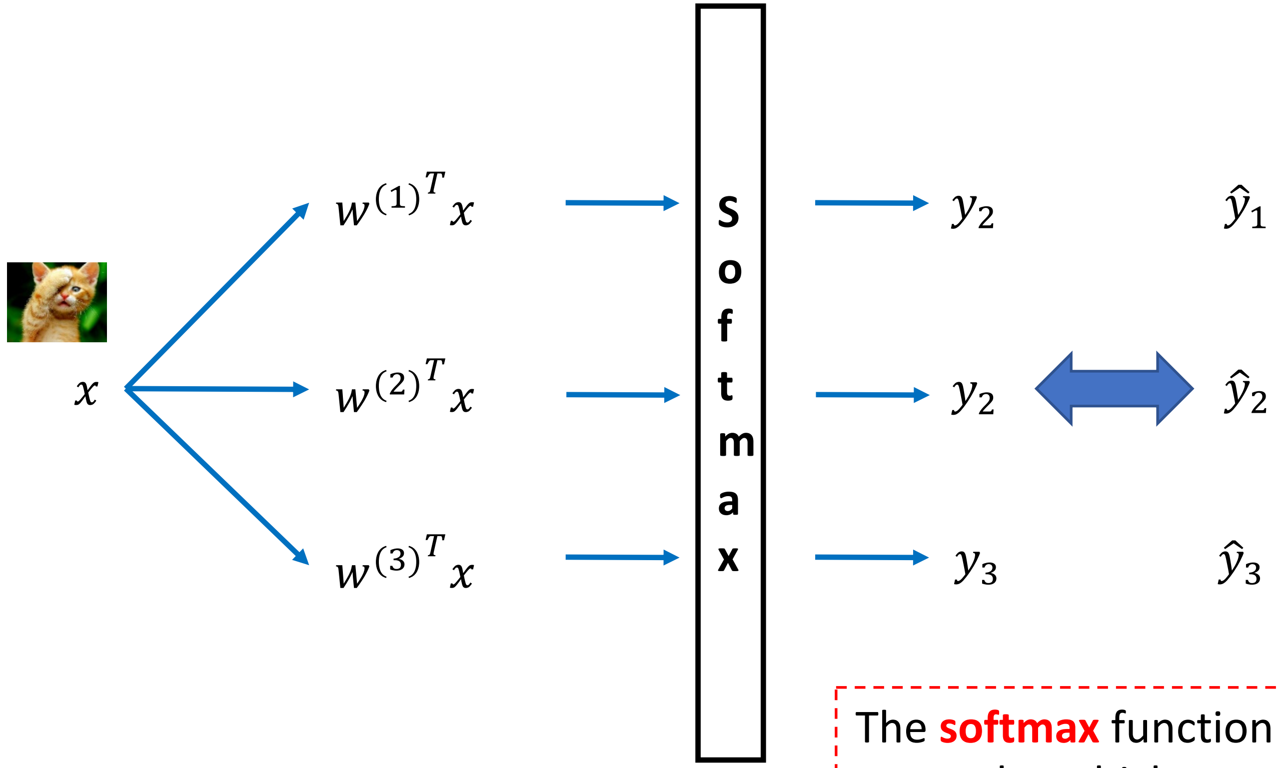
Softmax

Courtesy of Dr. Hung-yi Lee

Multi-Class Classification



- Consider multi classes at the same time



The **softmax** function is often used in the final layer of neural networks, which are applied to classification problems

Courtesy of Dr. Hung-yi Lee

CECS 456: Machine Learning (Spring 2020)

1. In softmax regression (Multinomial Logistic Regression), we compute the probability $P(y = k|x)$ for each value of $k = 1, 2, ..., K$ as

$$\begin{bmatrix} P(y = 1|x) \\ P(y = 2|x) \\ . \\ . \\ . \\ P(y = K|x) \end{bmatrix} = \frac{1}{\sum_{j=1}^{K} exp(\theta_j^T x)} \begin{bmatrix} exp(\theta_1^T x) \\ exp(\theta_2^T x) \\ . \\ . \\ . \\ exp(\theta_K^T x) \end{bmatrix},$$

where θ_j are the parameters.
Show that when $K = 2$, softmax regression reduces to the two-class logistic regression problem.