CS 66: Machine Learning

Prof. Sara Mathieson

Spring 2019

SWARTHMORE

Admin

- **Lab 3 due Thursday** (you should be close to done by lab on Wed)
- Midterm 1 Feb 27 (in lab)
  - Let me know ASAP about accommodations
- Office Hours **TODAY 12:30-2pm**
- Choose partner for Lab 4 by Tues night

Research Talk

Monday, February 18, 2019

11:30-12:30 p.m.

Science Center 240

Michael Wehar

Temple University

“Formal Language Theory and Applications”

Outline for February 18

- Intuition behind logistic regression
- Gradient descent for logistic regression
- Multi-class logistic regression

Soon: statistical evaluation

Outline for February 18

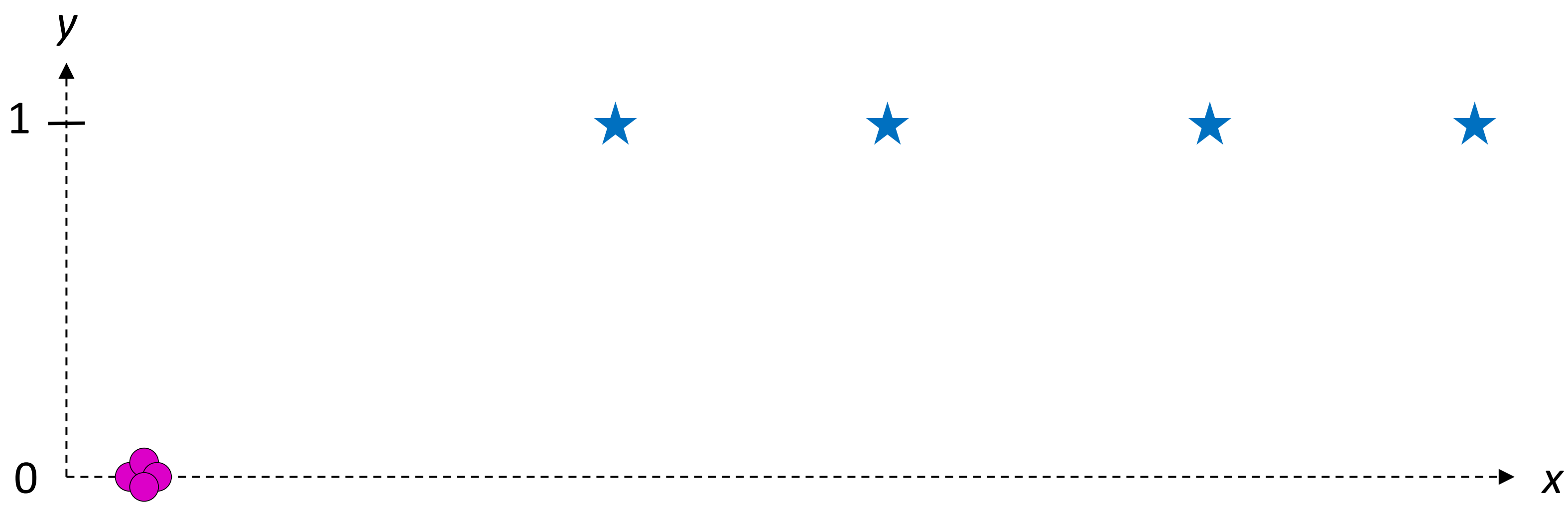
- Intuition behind logistic regression

- Gradient descent for logistic regression

- Multi-class logistic regression

Soon: statistical evaluation

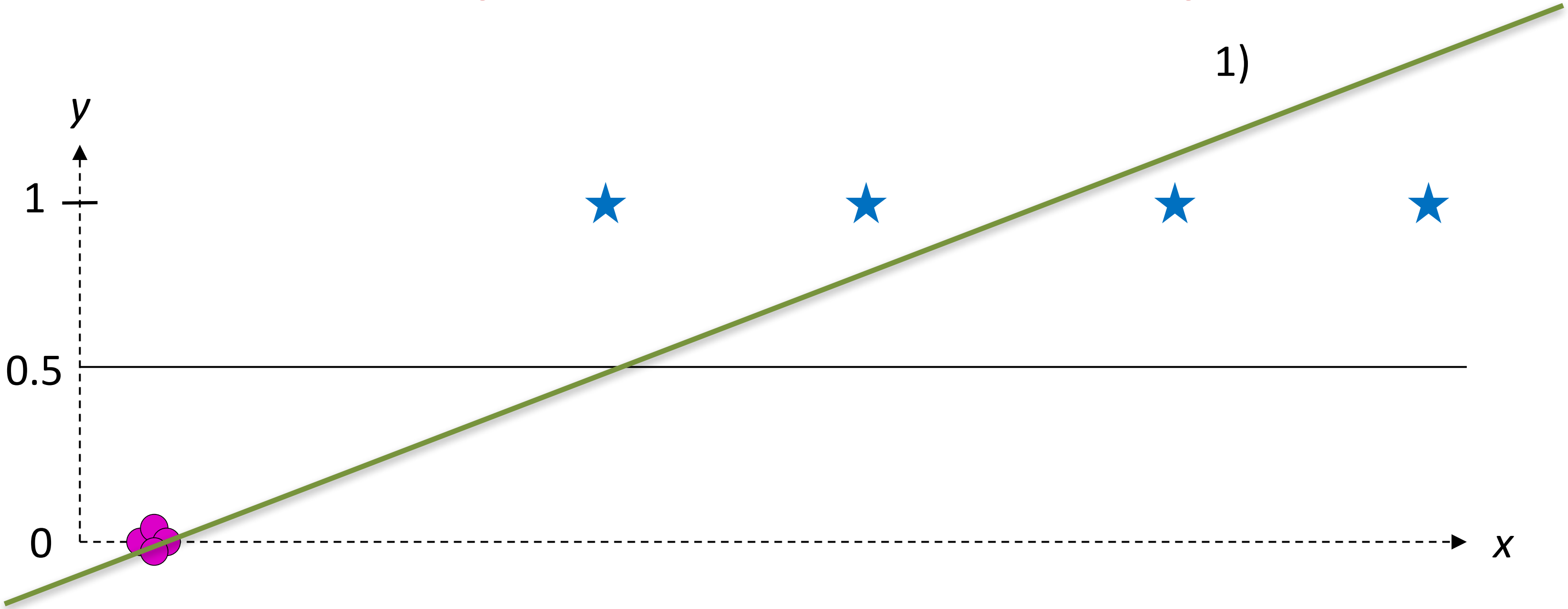
Informal quiz: discuss with a partner



1) What line of best fit would be produced by linear regression? (roughly)

2) What linear decision boundary would be produced by logistic regression? (roughly)

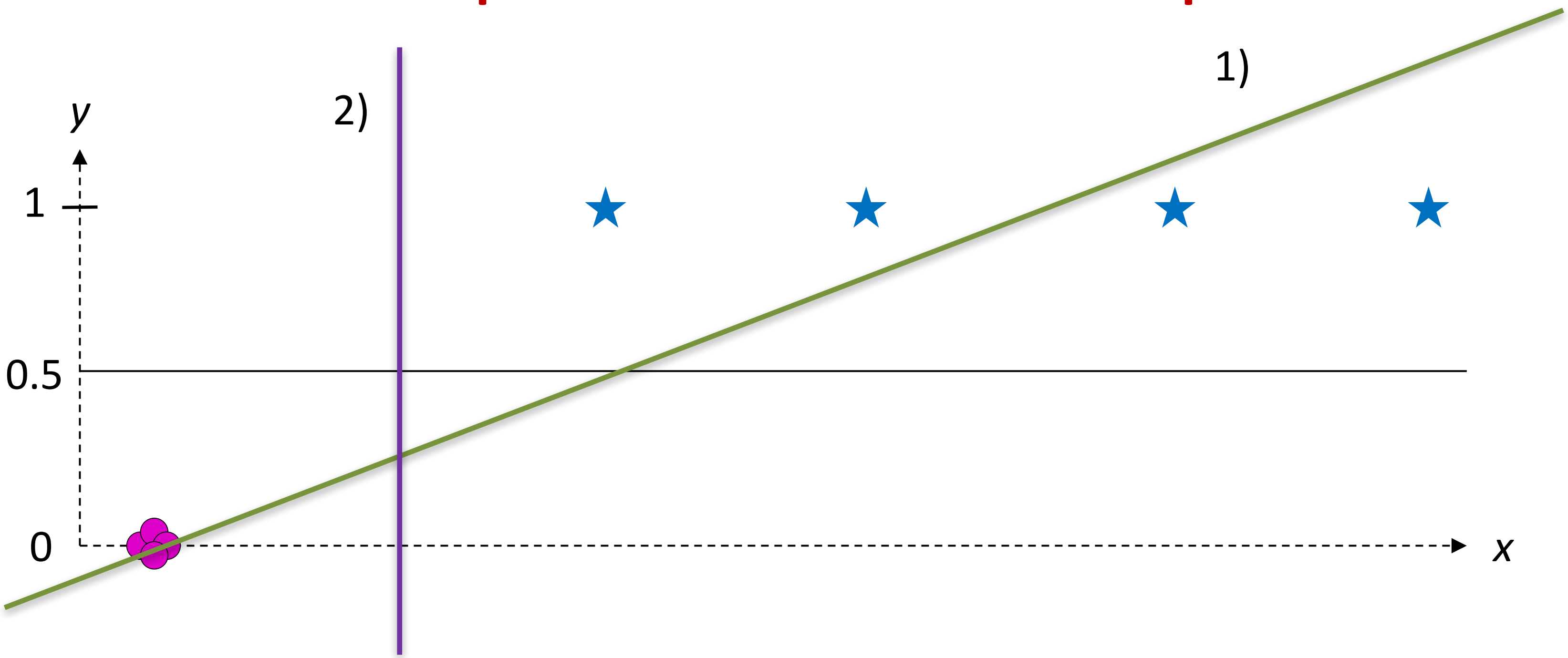
Informal quiz: discuss with a partner



1) What line of best fit would be produced by linear regression? (roughly)

2) What linear decision boundary would be produced by logistic regression? (roughly)

Informal quiz: discuss with a partner



1) What line of best fit would be produced by linear regression? (roughly)

2) What linear decision boundary would be produced by logistic regression? (roughly)

Outline for February 18

- Intuition behind logistic regression

- Gradient descent for logistic regression

- Multi-class logistic regression

Soon: statistical evaluation

Logistic Regression

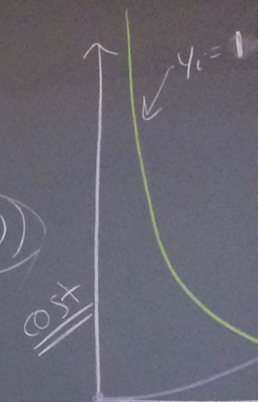
Likelihood function

minimize $\left\{ \begin{array}{l} \boxed{\text{cost}} \\ \hookrightarrow \underline{\text{negative}}\ \underline{\text{log}}\ \underline{\text{likelihood}} \end{array} \right.$

$$L(\vec{b}) = \prod_{i=1}^{n} \underbrace{\left(h_{\vec{b}}(\vec{x}_i)\right)^{y_i}}_{\text{prob } y_i=1} \underbrace{\left(1-h_b(\vec{x}_i)\right)^{1-y_i}}_{\text{prob } y_i=0}$$

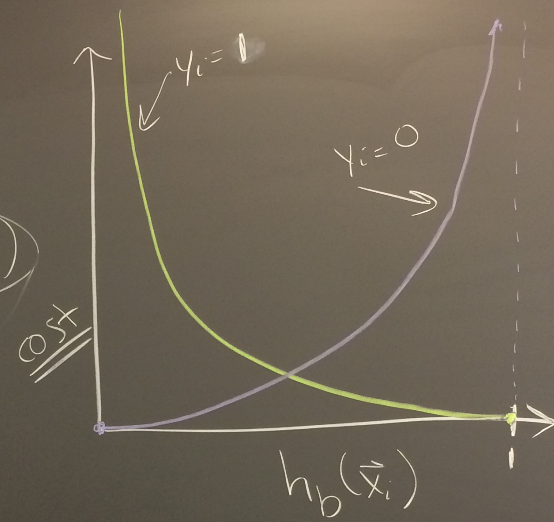
$$J(\vec{b}) = -\sum_{i=1}^{n} y_i \log h_b(\vec{x}_i) - \sum_{i=1}^{n} (1-y_i) \log\left(1-h_b(\vec{x}_i)\right)$$

$$\boxed{\text{cost}} \text{ (for 1 } x_i) = \begin{cases} -\log h_b(\vec{x}_i) & \text{if } y_i = 1 \\ -\log\left(1-h_b(\vec{x}_i)\right) & \text{if } y_i = 0 \end{cases}$$



y_i

$\sum (1-y_i) \log(1-h_b(\vec{x}_i))$

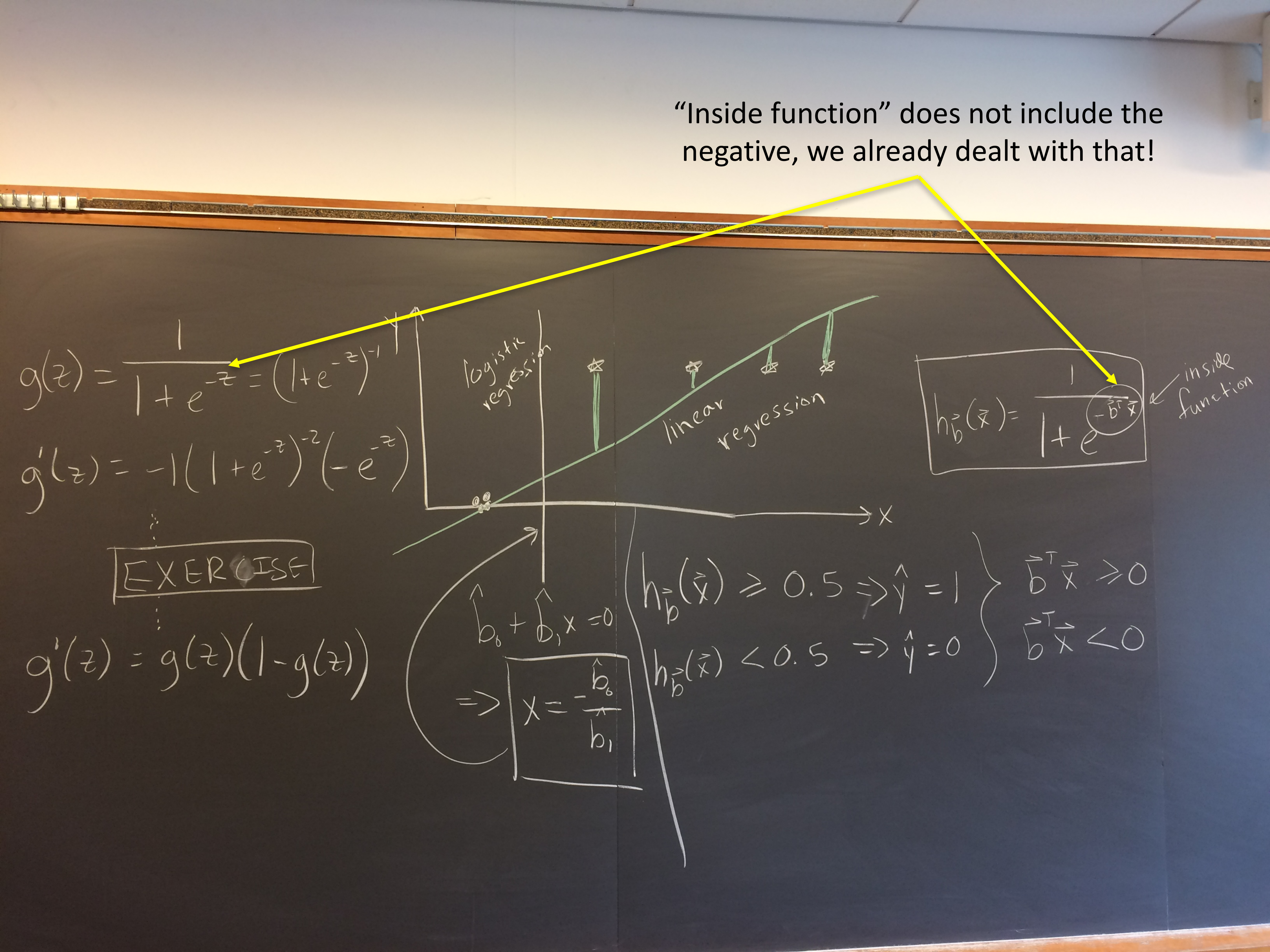


$\frac{\partial J}{\partial b_j}$

$y_i=1$

f $y_i=0$

"Inside function" does not include the negative, we already dealt with that!



$$\frac{\partial J}{\partial b_j} = \left[-\frac{y_i}{h_{\vec b}(\vec x_i)} + \frac{(1-y_i)}{1-h_{\vec b}(\vec x_i)}\right]\frac{\partial h_b(\vec x_i)}{\partial b_j}$$

$$= \left[-\frac{y_i}{h_{\vec b}(\vec x_i)} + \frac{(1-y_i)}{1-h_{\vec b}(\vec x_i)}\right] h_{\vec b}(\vec x_i)\left(1-h_{\vec b}(\vec x_i)\right)(+x_{ij})$$

$$= \left[-y_i\left(1-h_b(\vec x_i)\right) + (1-y_i)\left(h_{\vec b}(\vec x_i)\right)\right](+x_{ij})$$

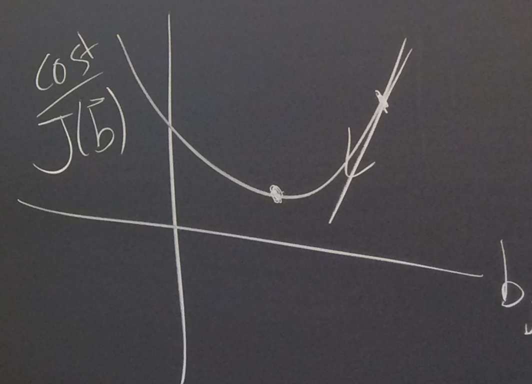
$$= \left[-y_i + \cancel{y_i h_b(\vec x_i)} + h_b(\vec x_i) - \cancel{y_i h_b(\vec x_i)}\right](+x_{ij})$$

$$\to \left(h_b(\vec x_i) - y_i\right)(+x_{ij})$$

$\boxed{\text{SGD}}$

same as linear regression!

$$\boxed{b_j \leftarrow b_j - \alpha\left(h_{\vec b}(\vec x_i) - y_i\right)x_{ij}}$$



Outline for February 18

- Intuition behind logistic regression
- Gradient descent for logistic regression
- Multi-class logistic regression

Soon: statistical evaluation

Multi-class Logistic Regression

- political party
- blood groups

K classes

$$y \in \{1, 2, \cdots, K\}$$

$$h_b(\vec{x}) = \frac{1}{1 + e^{-\vec{b}^T\vec{x}}}$$

$$h_b(\vec{x}) = \frac{e^{\vec{b}^T\vec{x}}}{\underbrace{e^{\vec{b}^T\vec{x}}}_{\substack{\text{weight on class}\\ y=1}} + \underbrace{1}_{\substack{\text{weight on class}\\ y=0}}}$$

$$\underbrace{1 - h_b(\vec{x})}_{\substack{\text{prob}\\ y=0}} = \frac{1}{e^{\vec{b}^T\vec{x}} + 1}$$

$$h_B(\vec{x}) = \underbrace{\begin{bmatrix} p(y=1|\vec{x}) \\ p(y=2|\vec{x}) \\ \vdots \\ p(y=K|\vec{x}) \end{bmatrix}}_{\text{must sum to 1}} = \underbrace{\frac{1}{\sum_{k=1}^{K} e^{\vec{b}^{(k)T}\vec{x}}}}_{\text{normalizer}} \begin{bmatrix} e^{\vec{b}^{(1)T}\vec{x}} \\ e^{\vec{b}^{(2)T}\vec{x}} \\ \vdots \\ e^{\vec{b}^{(K)T}\vec{x}} \end{bmatrix}$$

$$B = \underbrace{\begin{bmatrix} | & | & & | \\ \vec{b}^{(1)} & \vec{b}^{(2)} & \cdots & \vec{b}^{(K)} \\ | & | & & | \end{bmatrix}}_{(p+1)\times K}$$

cost function

$$J_B(\vec{x}) = -\sum_{i=1}^{n} \sum_{k=1}^{K} \underbrace{\mathbb{1}(y_i = k)}_{\text{indicator}} \underbrace{\log p(y_i = k \mid \vec{x}_i)}_{\text{cost of class } k}$$

if true => 1

else => 0

K