CS 66: Machine Learning

Prof. Sara Mathieson

Spring 2019

SWARTHMORE

Admin

- Office hours **TODAY** 12:30-2pm
- **Midterm 1 Feb 27** (in lab)
  - Scribe notes posted
- **Lab 4 due March 8** (Friday before spring break)

Outline for February 25

- Review
  - Lab 3 solutions
  - Logistic Regression

- Naïve Bayes for continuous features (if time)

Outline for February 25

- Review
  - Lab 3 solutions
  - Logistic Regression
- Naïve Bayes for continuous features (if time)

Review: overfitting

- Consider a hypothesis: *h*
  - Training error: $error_{train}(h)$
  - Error over all possible data: $error_D(h)$

- A hypothesis *h* **overfits** training data if there exists another hypothesis *h'* s.t.
  - $error_{train}(h) < error_{train}(h')$ AND
  - $error_D(h) > error_D(h')$

Lab 3 Solutions

- [not posted online]

Lab 3 vector math

$$J^R(b) = J(b) + \lambda \sum_{j=1}^{p} b_j^2$$

$$\boxed{\lambda = 10^{-5}}$$

$$X = \underbrace{\begin{bmatrix} 1 & — & x_1 & — \\ 1 & — & x_2 & — \\ 1 & — & x_3 & — \\ \vdots & & \vdots & \\ 1 & — & x_n & — \end{bmatrix}}_{n \times \cancel{(p+1)}} \underbrace{\begin{bmatrix} | \\ \vec{b} \\ | \end{bmatrix}}_{\cancel{(p+1)} \times 1} = \underbrace{\begin{bmatrix} | \\ \hat{y} \\ | \end{bmatrix}}_{n \times 1}$$

$$\vec{b}^T \vec{x} = \begin{bmatrix} — & b^T & — \end{bmatrix} \begin{bmatrix} | \\ \vec{x} \\ | \end{bmatrix}$$

Outline for February 25

- Review
  - Lab 3 solutions
  - Logistic Regression
- Naïve Bayes for continuous features (if time)

SGD for i in range(n):

linear & logistic

$$\underbrace{\begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_p \end{bmatrix}}_{\text{model}} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_p \end{bmatrix} - \alpha \underbrace{\left(h_{\vec{b}}(\vec{x}_i) - y_i \right)}_{\text{scalar}} \begin{bmatrix} x_{i0} \\ x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{bmatrix}$$

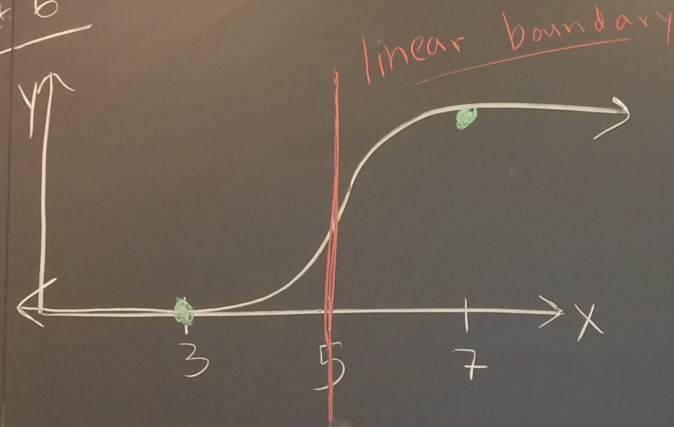
$$\boxed{h_{\vec{b}}(\vec{x}_i) = \frac{1}{1 + e^{-\vec{b}^T \vec{x}_i}}}$$

logistic

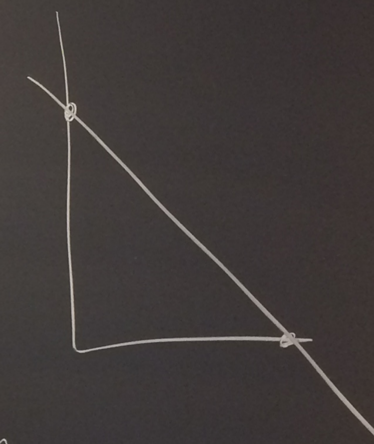
$$\boxed{h_{\vec{b}}(\vec{x}_i) = \vec{b}^T \vec{x}_i}$$

linear

Handout 6



likelihood function



$$L(\vec{b}) = \prod_{i=1}^{n} h_b(x_i)^{y_i}\left(1-h_b(x_i)\right)^{1-y_i}$$

no closed form solution

$$\boxed{L(p) = \prod_{i=1}^{n} p^{y_i}(1-p)^{1-y_i}}$$

coin flip

(b)

$$L(\vec{b}) = \underbrace{\left(1 - h_{\vec{b}}(3)\right)}_{x_1} \underbrace{h_{\vec{b}}(7)}_{x_2}$$

(c)

$$-\log L(\vec{b}) = \underbrace{J(\vec{b})}_{\text{cost}}$$

$\boxed{\text{minimize}}$

(d)

1 took derivative wrt b_j

2 use derivative for one x_i in SGD

(c)

$$\begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} \cancel{b_0}^{\,0} \\ \cancel{b_1}^{\,0} \end{bmatrix} - 0.1\left(\frac{1}{1+e^0} - 0\right)\begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

$$= \begin{bmatrix} -0.1\left(\frac{1}{2}\right)\cdot 1 \\ -0.1\left(\frac{1}{2}\right)\cdot 3 \end{bmatrix} = \begin{bmatrix} -0.05 \\ -0.15 \end{bmatrix}$$

$$\underbrace{-5 + x}_{} = 0 \qquad x = 5$$

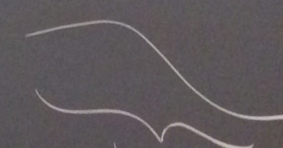
$$\boxed{\begin{matrix} b_0 = -5 \\ b_1 = 1 \end{matrix}}$$

(d)

$$\begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} 0.05 \\ 0.35 \end{bmatrix}$$

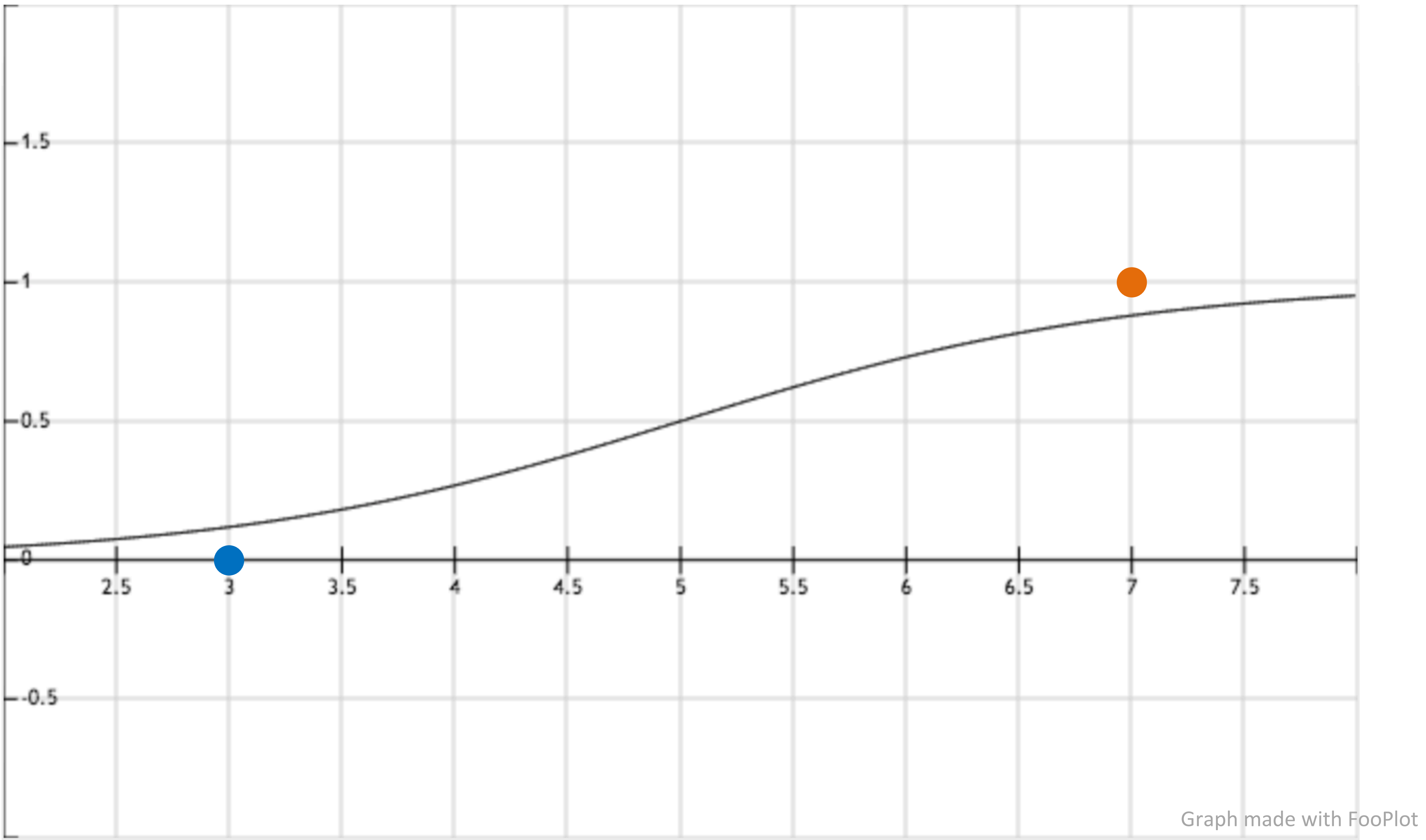
$$-10 + 2x = 0 \Rightarrow \boxed{x = 5}$$

$$\boxed{\begin{matrix} b_0 = -10 \\ b_1 = 2 \end{matrix}}$$

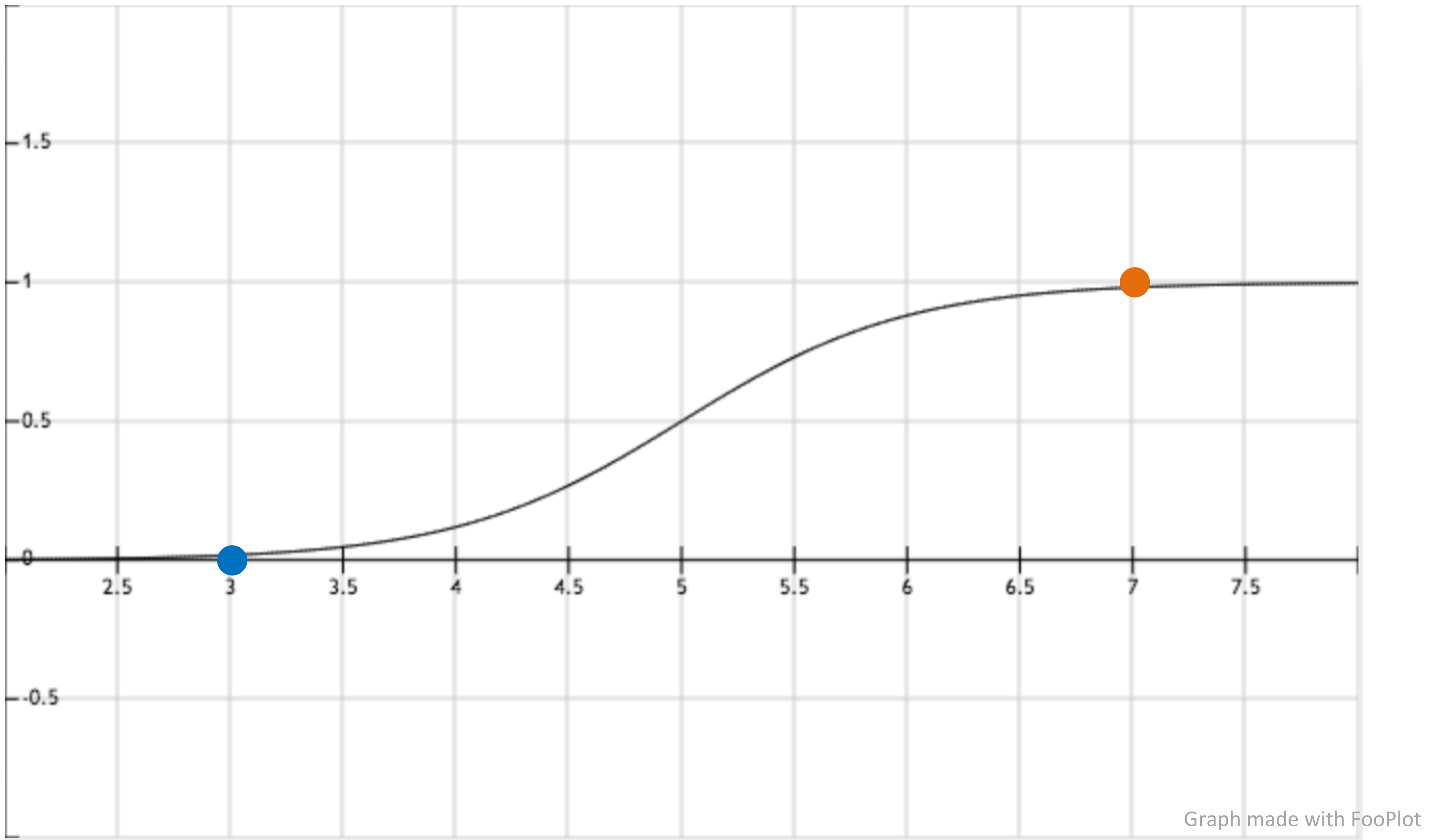


$b_1 < 0$

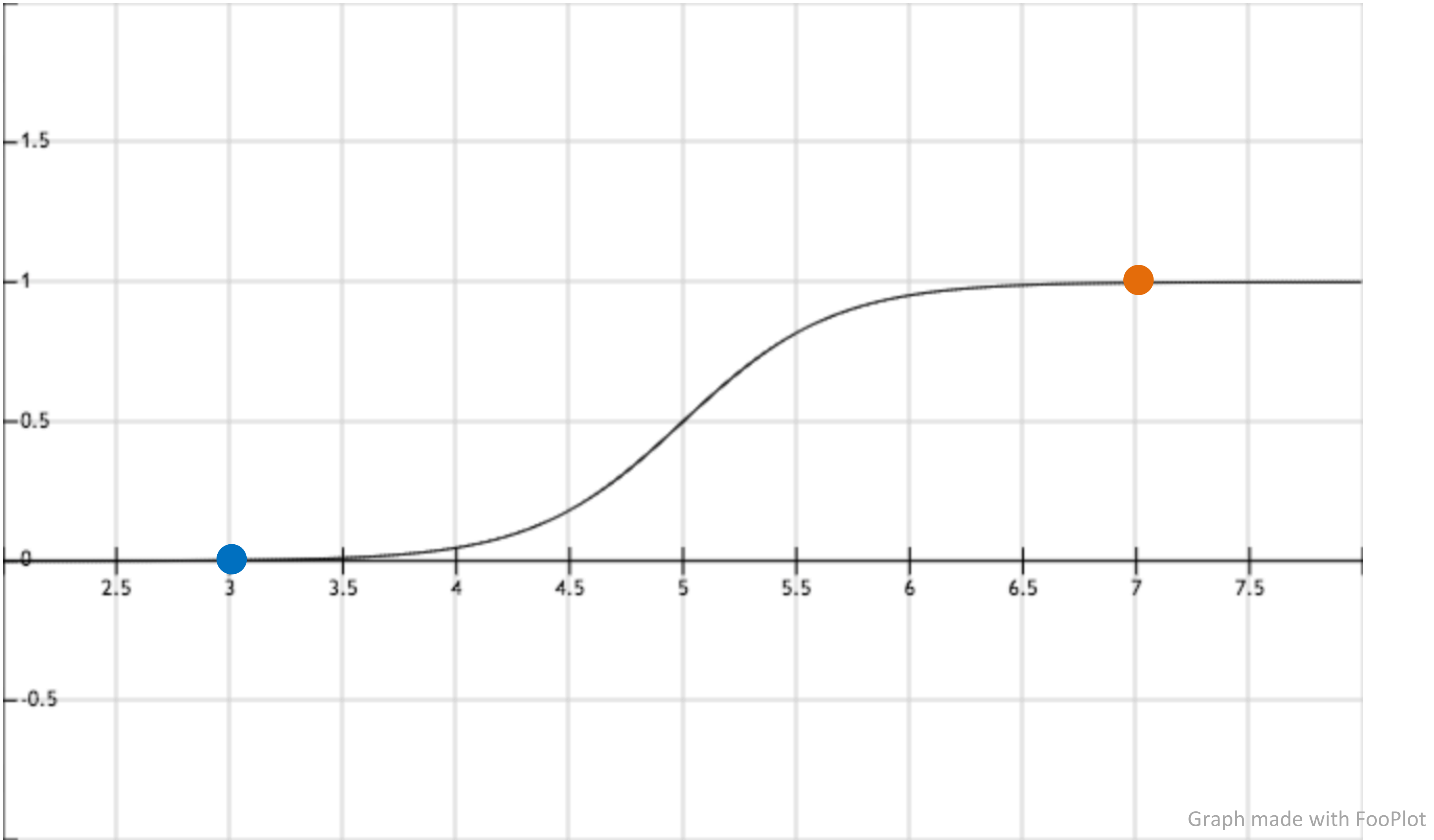
b_0=-5, b_1=1 $h_b(x) = 1/(1+e^{\wedge}(5-x))$



b_0=-10, b_1=2 $h_b(x) = 1/(1+e\^(10-2x))$



b_0=-15, b_1=3 $h_b(x) = 1/(1+e^{(15-3x)})$



b_0=-100, b_1=20 $h_b(x) = 1/(1+e^\wedge(100-20x))$