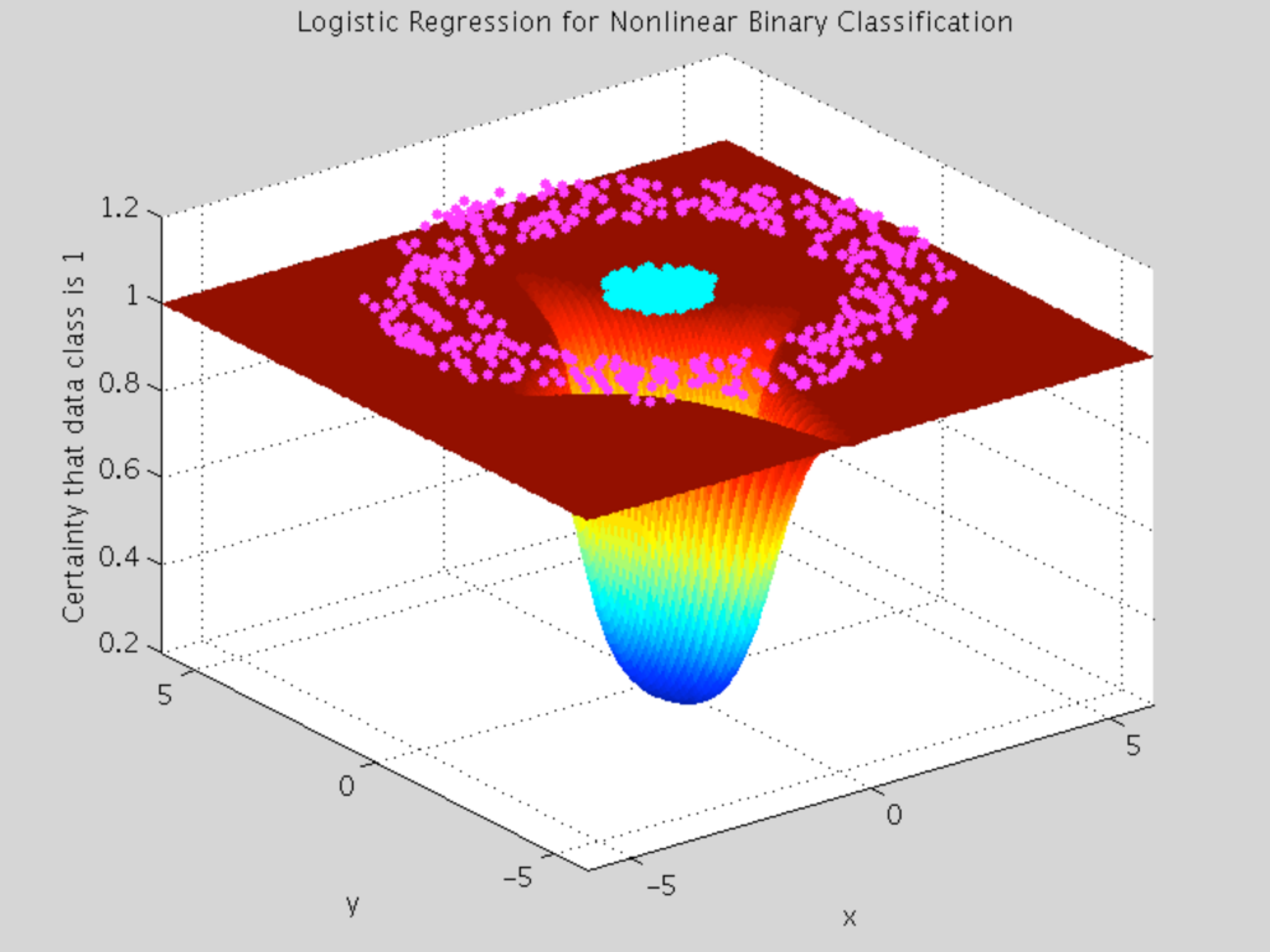
Logistic Regression For Programmers



Logistic Regression for Nonlinear Binary Classification
Certainty that data class is 1
1.2
1
0.8
0.6
0.4
0.2
5
0
−5
y
5
0
−5
x

$$\mathbf{X} = \begin{bmatrix} x_{0,0} & x_{0,1} & \cdots & x_{0,n} \\ x_{1,0} & x_{1,1} & \cdots & x_{1,n} \\ \cdots & \cdots & \ddots & \vdots \\ x_{m,0} & x_{m,1} & \cdots & x_{m,n} \end{bmatrix}$$

$$\mathbf{X} = \begin{bmatrix} x_{0,0} & x_{0,1} & \cdots & x_{0,n} \\ x_{1,0} & x_{1,1} & \cdots & x_{1,n} \\ \cdots & \cdots & \ddots & \vdots \\ x_{m,0} & x_{m,1} & \cdots & x_{m,n} \end{bmatrix}$$

$$\mathbf{y} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_m \end{bmatrix}$$

$$\mathbf{X} = \begin{bmatrix} x_{0,0} & x_{0,1} & \cdots & x_{0,n} \\ x_{1,0} & x_{1,1} & \cdots & x_{1,n} \\ \cdots & \cdots & \ddots & \vdots \\ x_{m,0} & x_{m,1} & \cdots & x_{m,n} \end{bmatrix}$$

$$\mathbf{y} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_m \end{bmatrix}$$

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_0 & \theta_1 & \cdots & \theta_n \end{bmatrix}$$

$$\mathbf{X} = \begin{bmatrix} x_{0,0} & x_{0,1} & \cdots & x_{0,n} \\ x_{1,0} & x_{1,1} & \cdots & x_{1,n} \\ \cdots & \cdots & \ddots & \vdots \\ x_{m,0} & x_{m,1} & \cdots & x_{m,n} \end{bmatrix}$$

$$\mathbf{y} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_m \end{bmatrix}$$

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_0 & \theta_1 & \cdots & \theta_n \end{bmatrix}$$

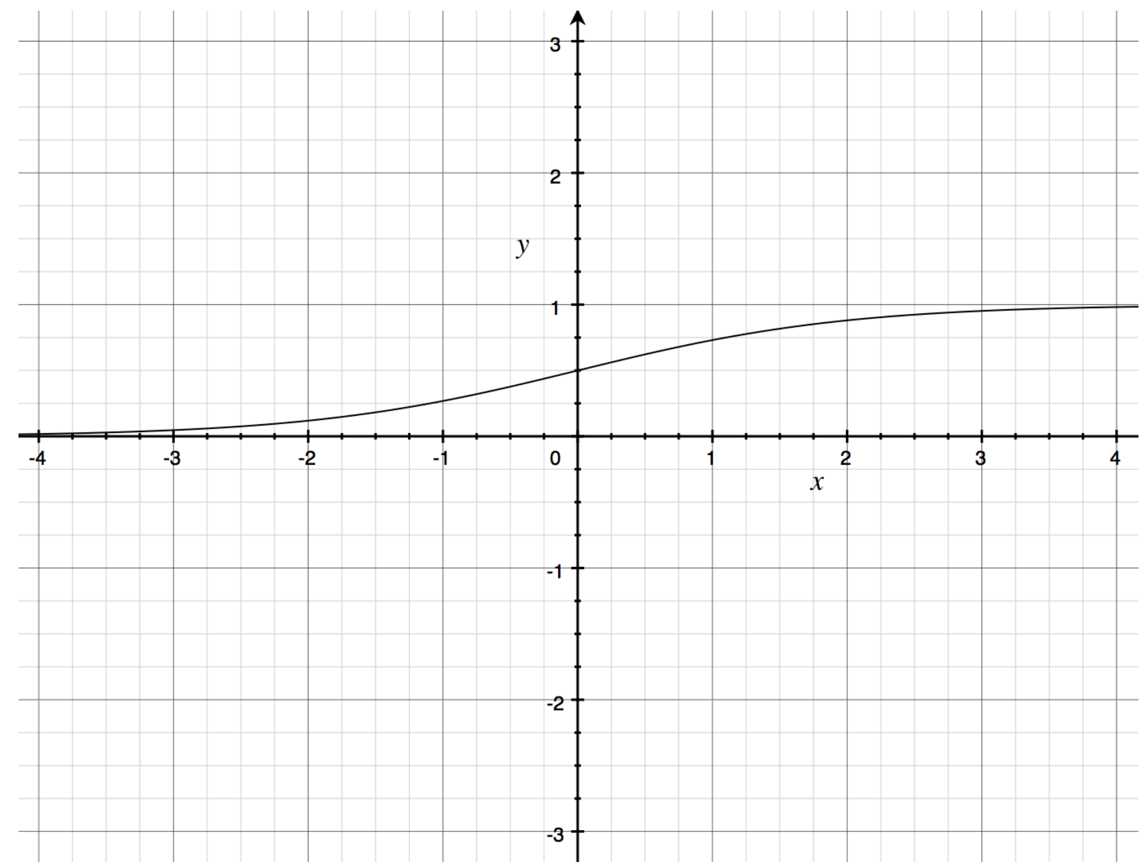
$$h(\mathbf{X}) = \frac{1}{1 + e^{-\boldsymbol{\theta}\mathbf{X}^{\top}}} = \mathbf{a} = \begin{bmatrix} a_0 & a_1 & \cdots & a_n \end{bmatrix}$$

$$\mathbf{X} = \begin{bmatrix} x_{0,0} & x_{0,1} & \cdots & x_{0,n} \\ x_{1,0} & x_{1,1} & \cdots & x_{1,n} \\ \cdots & \cdots & \ddots & \vdots \\ x_{m,0} & x_{m,1} & \cdots & x_{m,n} \end{bmatrix}$$

$$\mathbf{y} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_m \end{bmatrix}$$

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_0 & \theta_1 & \cdots & \theta_n \end{bmatrix}$$

$$h(\mathbf{X}) = \frac{1}{1+e^{-\boldsymbol{\theta}\mathbf{X}^{\top}}} = \mathbf{a} = \begin{bmatrix} a_0 & a_1 & \cdots & a_n \end{bmatrix}$$



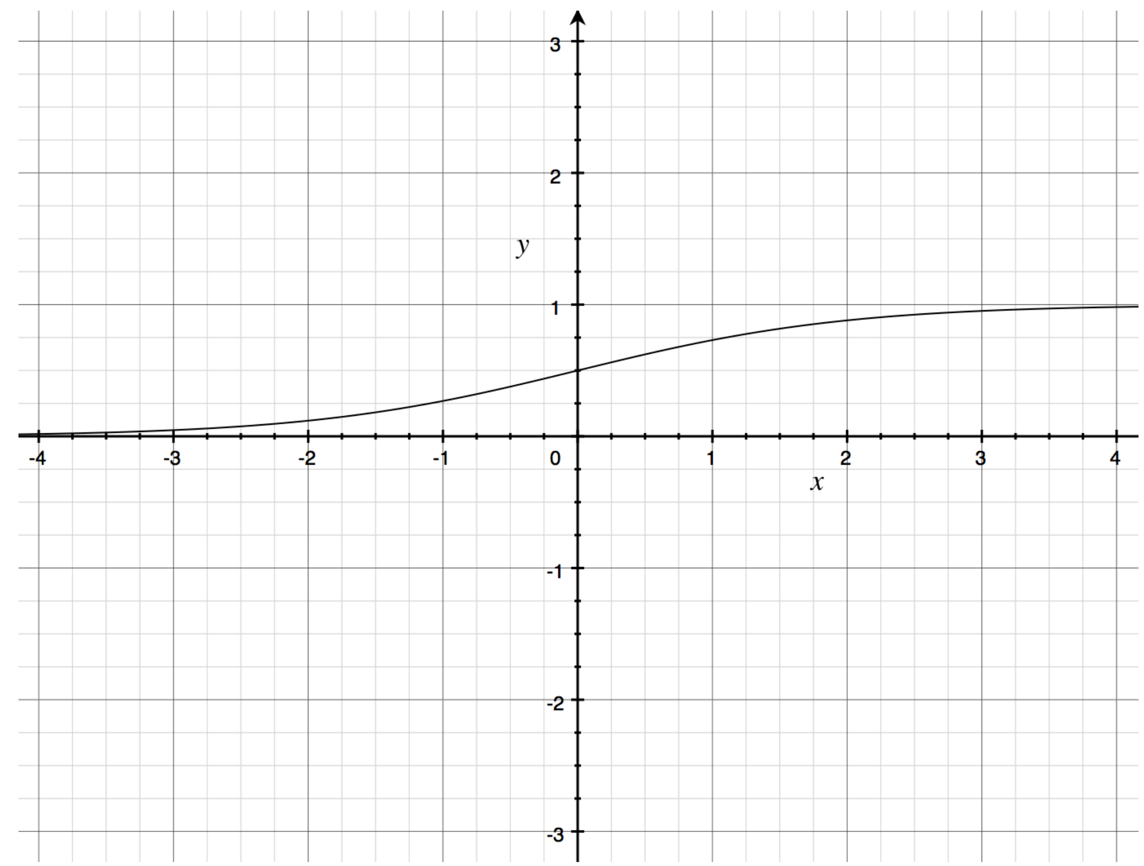
$$\mathbf{X} = \begin{bmatrix} x_{0,0} & x_{0,1} & \cdots & x_{0,n} \\ x_{1,0} & x_{1,1} & \cdots & x_{1,n} \\ \cdots & \cdots & \ddots & \vdots \\ x_{m,0} & x_{m,1} & \cdots & x_{m,n} \end{bmatrix}$$

$$\mathbf{y} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_m \end{bmatrix}$$

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_0 & \theta_1 & \cdots & \theta_n \end{bmatrix}$$

$$h(\mathbf{X}) = \frac{1}{1+e^{-\boldsymbol{\theta}\mathbf{X}^{\top}}} = \mathbf{a} = \begin{bmatrix} a_0 & a_1 & \cdots & a_n \end{bmatrix}$$

$$C = -\frac{1}{m}\left[\sum_{i=1}^{m} y_i \log\left(h(x_i)\right) + (1-y_i)\log\left(1-h(x_i)\right)\right]$$



$$\mathbf{X} = \begin{bmatrix} x_{0,0} & x_{0,1} & \cdots & x_{0,n} \\ x_{1,0} & x_{1,1} & \cdots & x_{1,n} \\ \cdots & \cdots & \ddots & \vdots \\ x_{m,0} & x_{m,1} & \cdots & x_{m,n} \end{bmatrix}$$

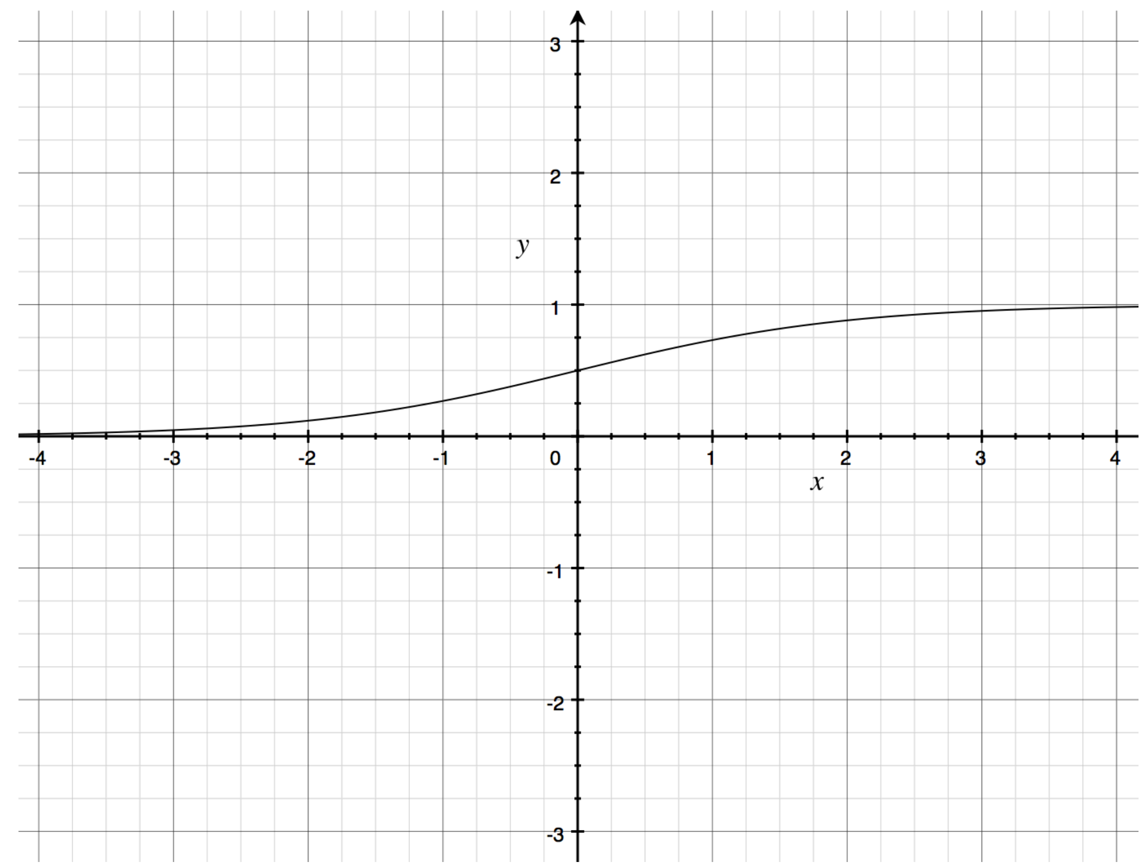
$$\mathbf{y} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_m \end{bmatrix}$$

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_0 & \theta_1 & \cdots & \theta_n \end{bmatrix}$$

$$h(\mathbf{X}) = \frac{1}{1+e^{-\boldsymbol{\theta}\mathbf{X}^{\top}}} = \mathbf{a} = \begin{bmatrix} a_0 & a_1 & \cdots & a_n \end{bmatrix}$$

$$C = -\frac{1}{m}\left[\sum_{i=1}^{m} y_i \log\left(h(x_i)\right) + (1-y_i)\log\left(1-h(x_i)\right)\right]$$

$$\frac{\partial C}{\partial \boldsymbol{\theta}} = \nabla_{\boldsymbol{\theta}} C = \frac{1}{m}\left(y^{\top} - h(\mathbf{X})\right)\mathbf{X} = \Delta\begin{bmatrix} \theta_0 & \theta_1 & \cdots & \theta_n \end{bmatrix}$$



$$\mathbf{X} = \begin{bmatrix} x_{0,0} & x_{0,1} & \cdots & x_{0,n} \\ x_{1,0} & x_{1,1} & \cdots & x_{1,n} \\ \cdots & \cdots & \ddots & \vdots \\ x_{m,0} & x_{m,1} & \cdots & x_{m,n} \end{bmatrix}$$

$$\mathbf{y} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_m \end{bmatrix}$$

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_0 & \theta_1 & \cdots & \theta_n \end{bmatrix}$$

$$h(\mathbf{X}) = \frac{1}{1+e^{-\boldsymbol{\theta}\mathbf{X}^\top}} = \mathbf{a} = \begin{bmatrix} a_0 & a_1 & \cdots & a_n \end{bmatrix}$$

$$C = -\frac{1}{m}\left[\sum_{i=1}^{m} y_i \log\big(h(x_i)\big) + (1-y_i)\log\big(1-h(x_i)\big)\right]$$

$$\frac{\partial C}{\partial \boldsymbol{\theta}} = \nabla_{\boldsymbol{\theta}} C = \frac{1}{m}\big(y^\top - h(\mathbf{X})\big)\mathbf{X} = \Delta\begin{bmatrix} \theta_0 & \theta_1 & \cdots & \theta_n \end{bmatrix}$$

$$\boldsymbol{\theta} := \boldsymbol{\theta} + \alpha\nabla_{\boldsymbol{\theta}} C$$