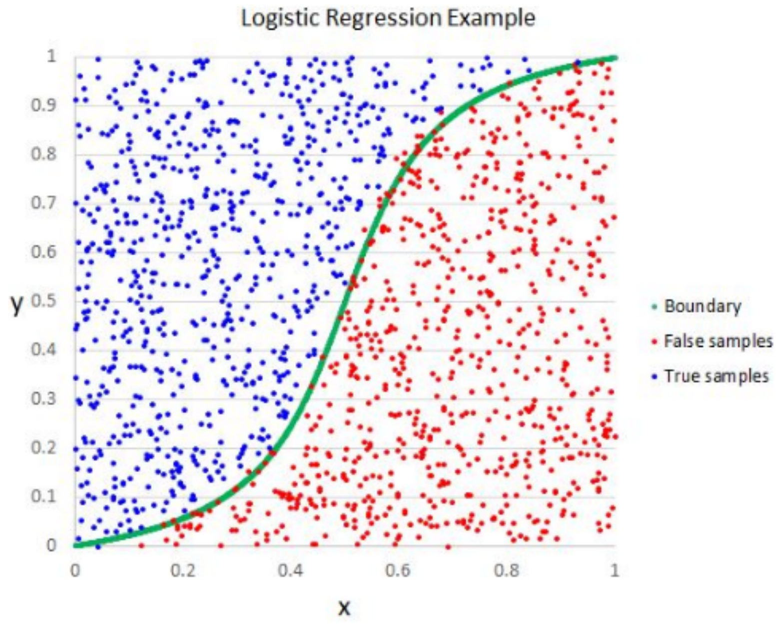
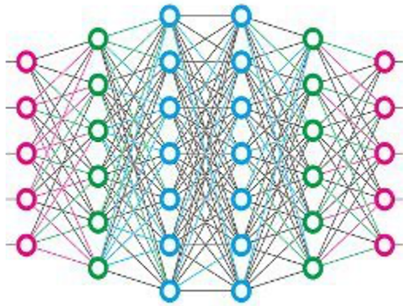
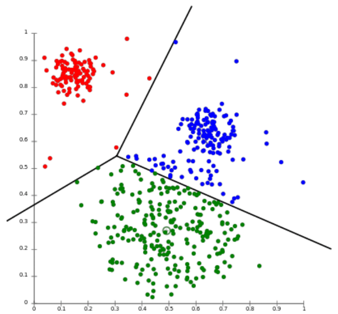
CS 445
Introduction to Machine Learning

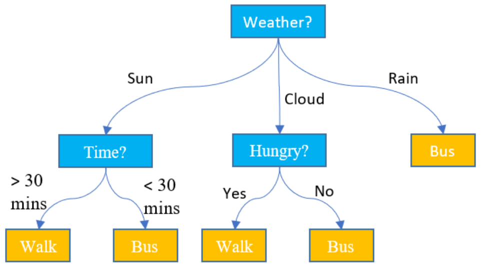
Logistic Regression

Instructor: Dr. Kevin Molloy









Review

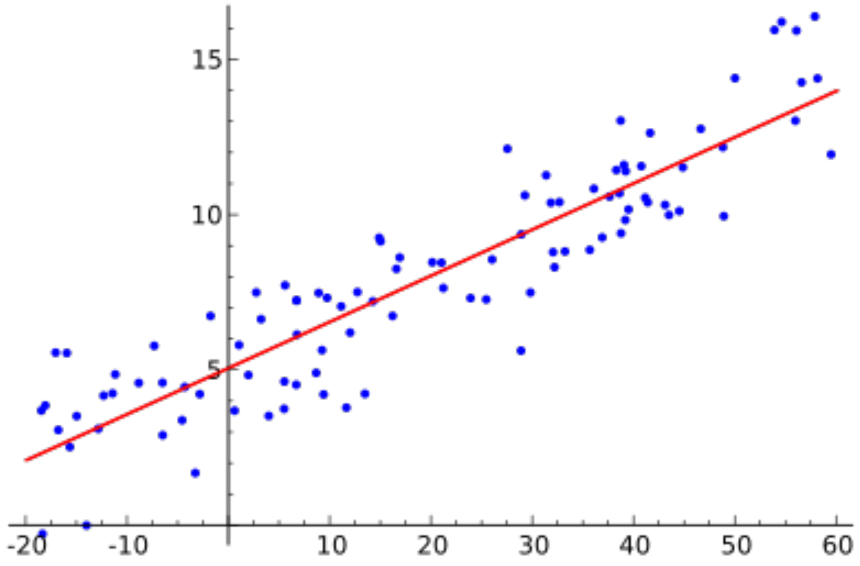
Linear regression

Finding the weights to assign to a polynomial so

that the resulting line minimizes the "loss".

$$h(x_1, x_2, \ldots x_n) = w_0 + w_1 x_1 + \ldots + w_n x_n$$

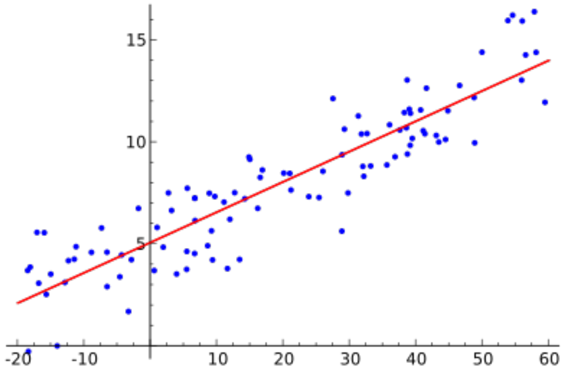
$$h(x) = w^T x$$



This function h(x) (hypothesis function) makes a real valued prediction (regression).

Linear Regression $L(w) = \frac{1}{2} \sum_{(x_i, y_i) \in D} (y_i - w^t x_i)^2$

Approach for Linear Regression



Linear Regression $L(w) = \frac{1}{2N} \sum_{(x_i, y_i) \in D} (y_i - w^t x_i)^2$

Optimize (find the min) of the loss function using the derivatives:

$$\frac{\partial L(w)}{\partial w_i} = \frac{1}{N} \sum_{j=0..N} x_j^{(i)} (y_j - w^t x_j)$$

$$\frac{\partial L(w)}{\partial w_0} = \frac{1}{N} \sum_{j=0..N} (y_j - w^t x_j)$$

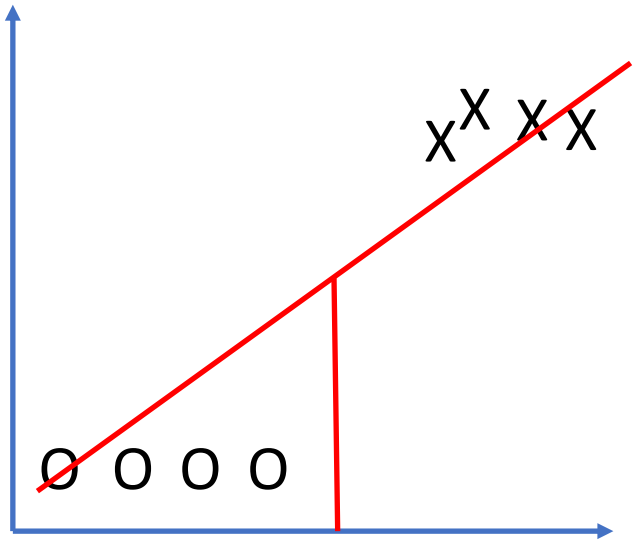
Linear Regression Algorithm

1. Make predictions using current w and compute loss
2. Compute derivative and update w's
3. When loss change is a little STOP, otherwise, go back to 1.

Logistic Regression

World's **WORST** algorithm name

Transform linear regression into a classification algorithm



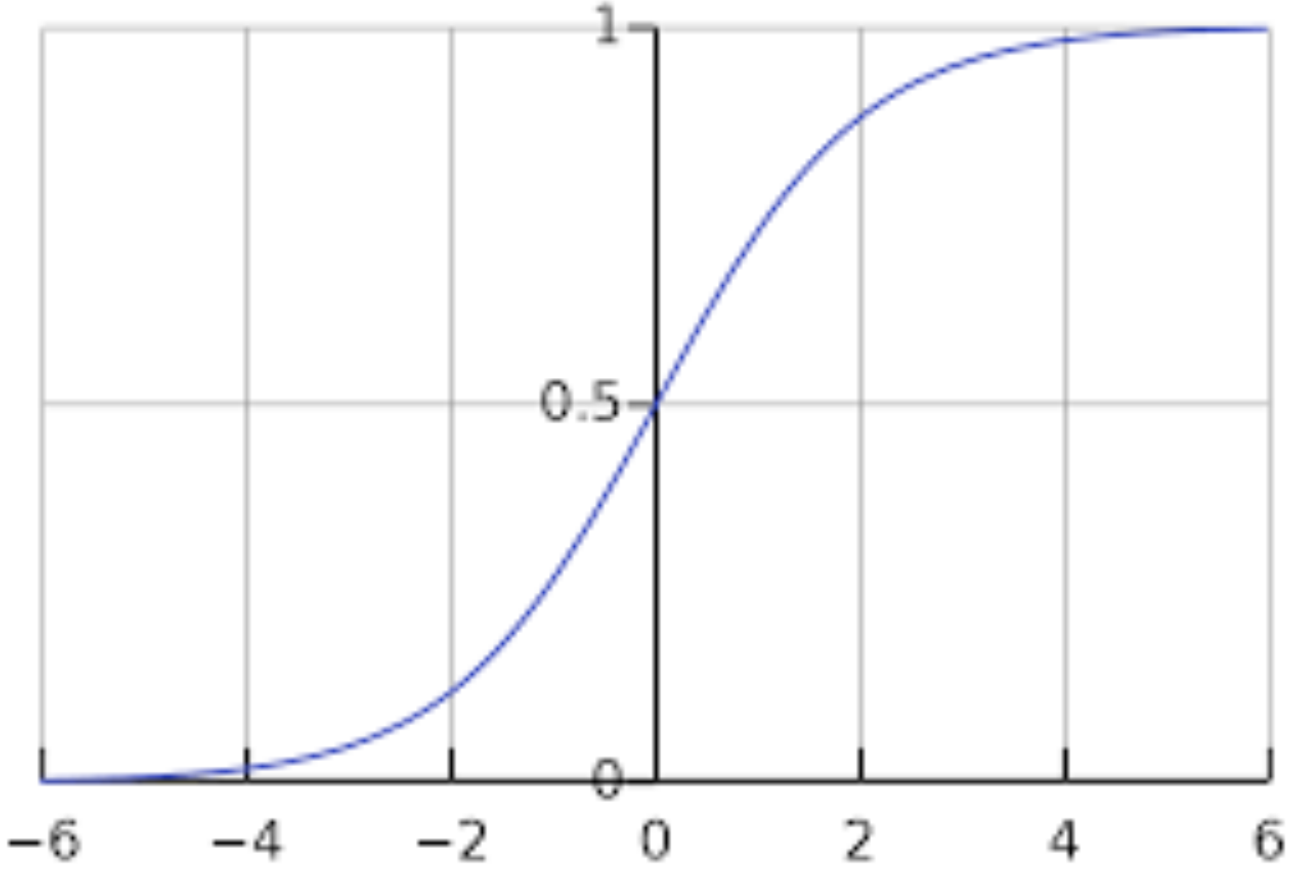
h(x) >= 0.5, predict y = 1 (X class)

h(x) < 0.5, predict y = 0 () class)

Map Function to Values Between 0 and 1

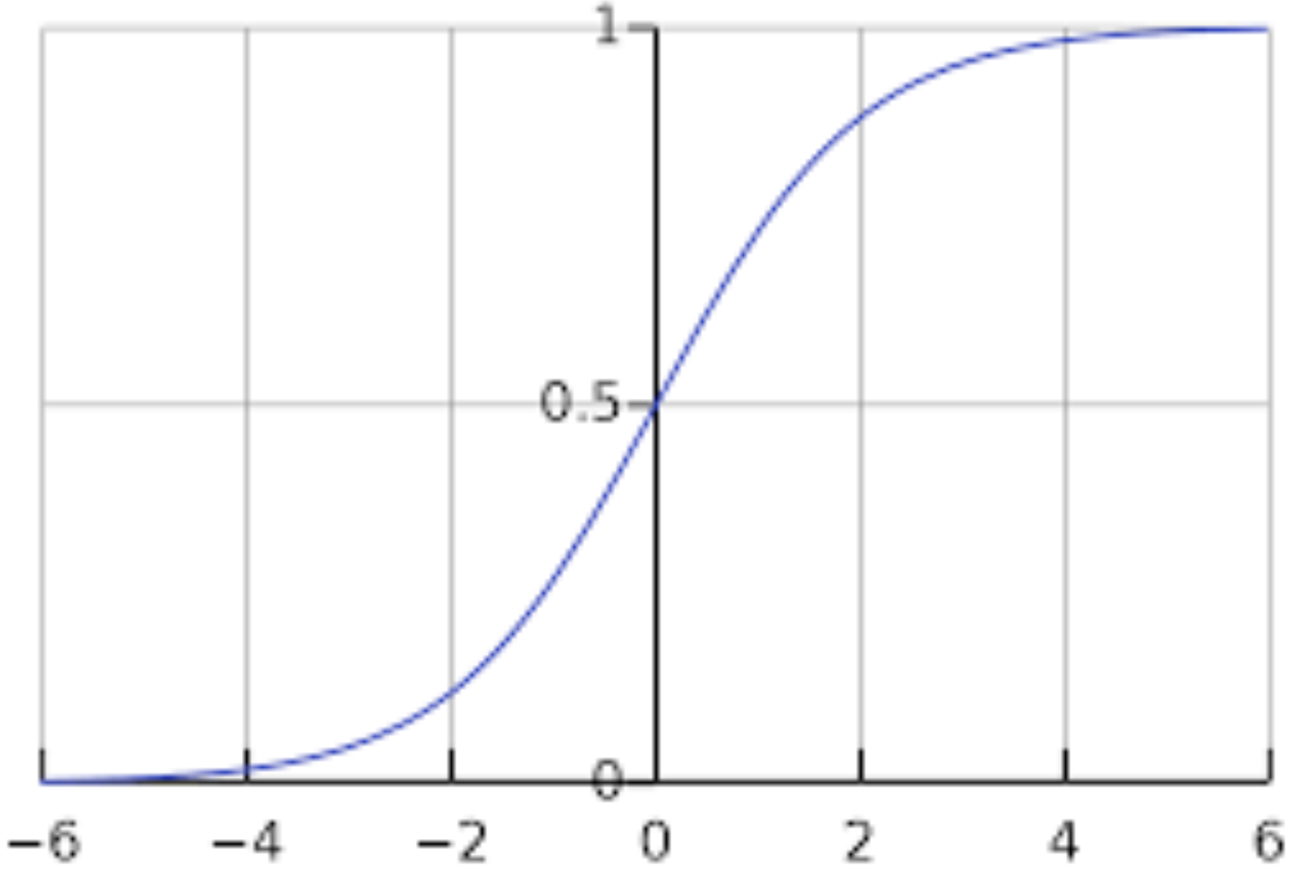
Sigmoid (z) = $\frac{1}{1+ e^{-z}}$

$$\frac{1}{1+\ e^{-w^t x}}$$



Different Loss Function

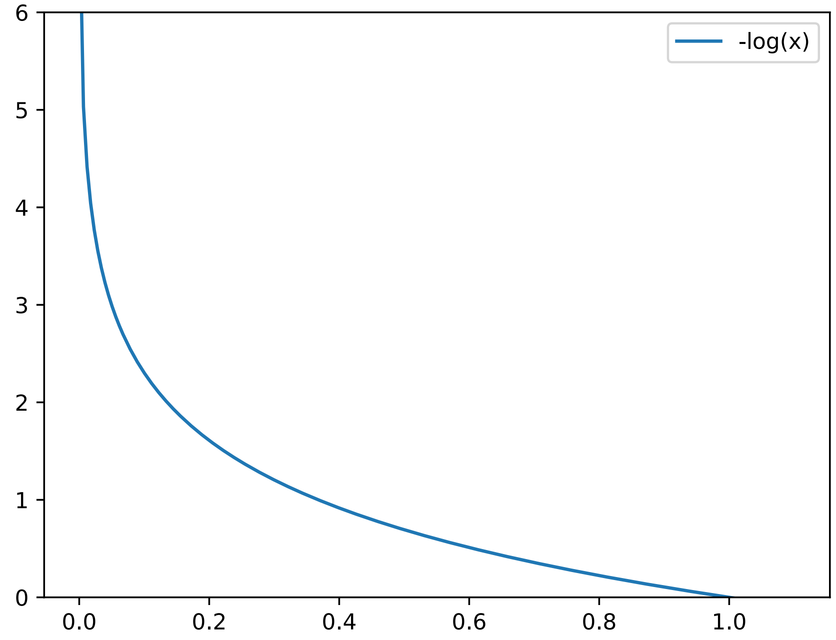
$$\frac{1}{1+\ e^{-w^t x}}$$

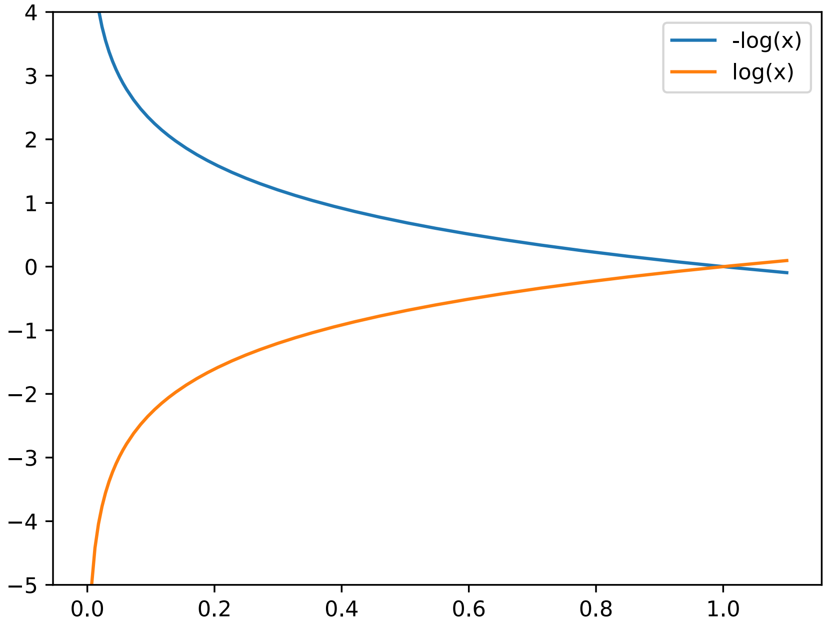


Linear Regression $L(w) = \frac{1}{2N}\sum_{(x_i, y_i)\in D}(y_i - w^t x_i)^2$

Cost Function for Linear Regression

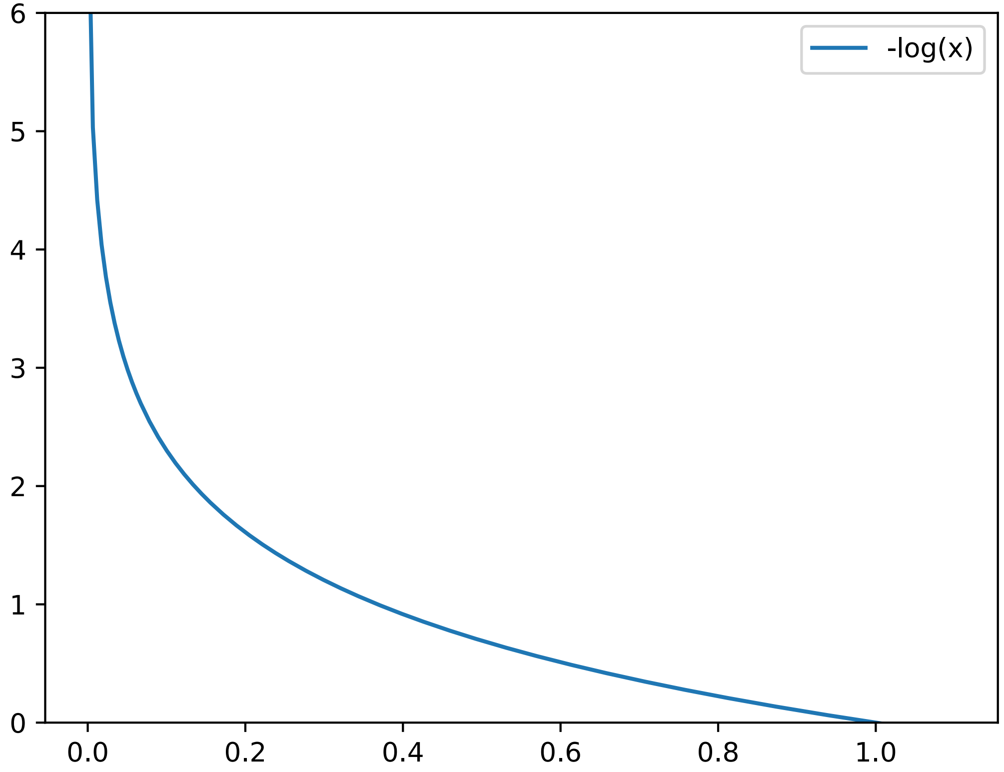
$$\text{Loss}(h(x), y) = \begin{cases} -\log\left(f_w(x)\right) & if\ y = 1 \\ -\log\left(1 - f_w(x)\right) & if\ y = 0 \end{cases}$$





Cost Function for Linear Regression

Loss($h(x)$, y) = $\begin{Bmatrix} -\log\left(f_w(x)\right) & if\ y = 1 \\ -\log\left(1 - f_w(x)\right) & if\ y = 0 \end{Bmatrix}$



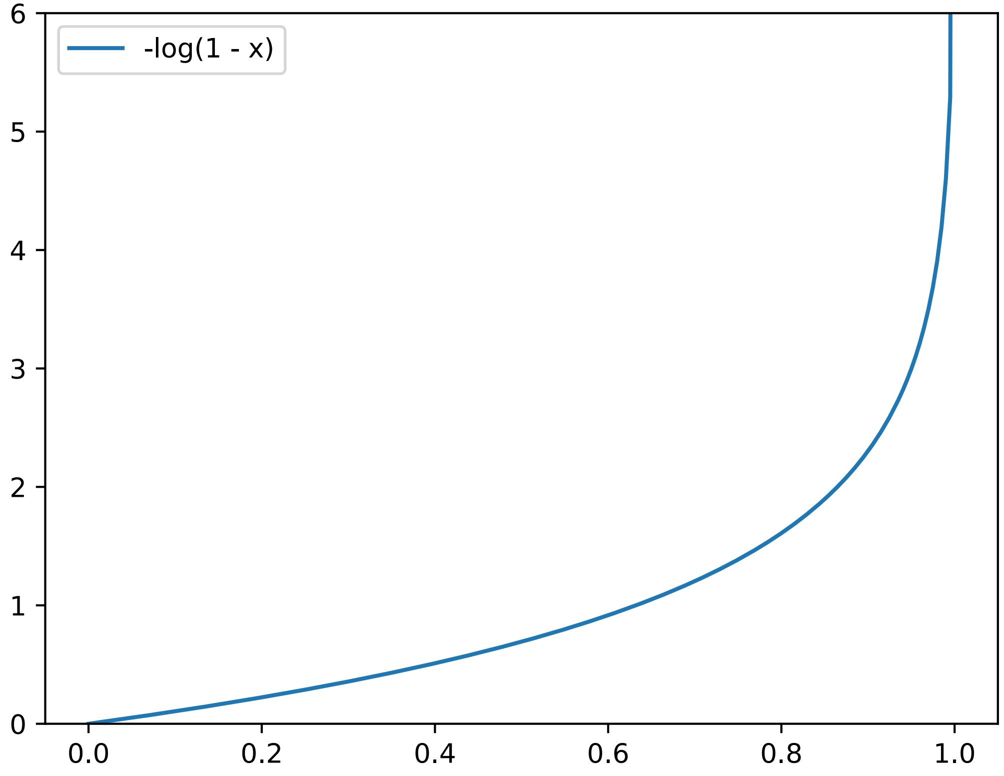
When y = 1:

f(x) = 1, then Cost = 0 (since (-log(1) = 0)

f(x) = 0 , then the loss (or penalty) will be very large.

Cost Function for Linear Regression

$$\text{Loss}(h(x), y) = \begin{cases} -\log\left(f_w(x)\right) & if\ y = 1 \\ -\log\left(1 - f_w(x)\right) & if\ y = 0 \end{cases}$$



When y = 0:

f(x) = 0, then Cost = 0 (since (-log(1 –f(x)) = 0)

f(x) = , then the loss (or penalty) will be very large.

Logistic Regression Loss

Loss(h(x), y) = $\begin{cases} -\log\left(f_w(x)\right) & if\ y = 1 \\ -\log\left(1 - f_w(x)\right) & if\ y = 0 \end{cases}$

Loss(h(x), y) = $\prod_{i=1}^{n} P(y = 1 \mid x_i)^{y_i} \;\; x \;\; P(y = 0 \mid x_i)^{1 - y_i}$