COSI 104a Introduction to machine learning

Chapter 6 – Logistic Regression

Instructor: Dr. Hongfu Liu
Email: hongfuliu@brandeis.edu

Main source: https://www-users.cs.umn.edu/~kumar001/dmbook/index.php

An Example

A binary variable

Age	Y disease
22	0
23	0
24	0
27	0
28	0
29	0
30	0
32	0
33	0
35	0
40	1

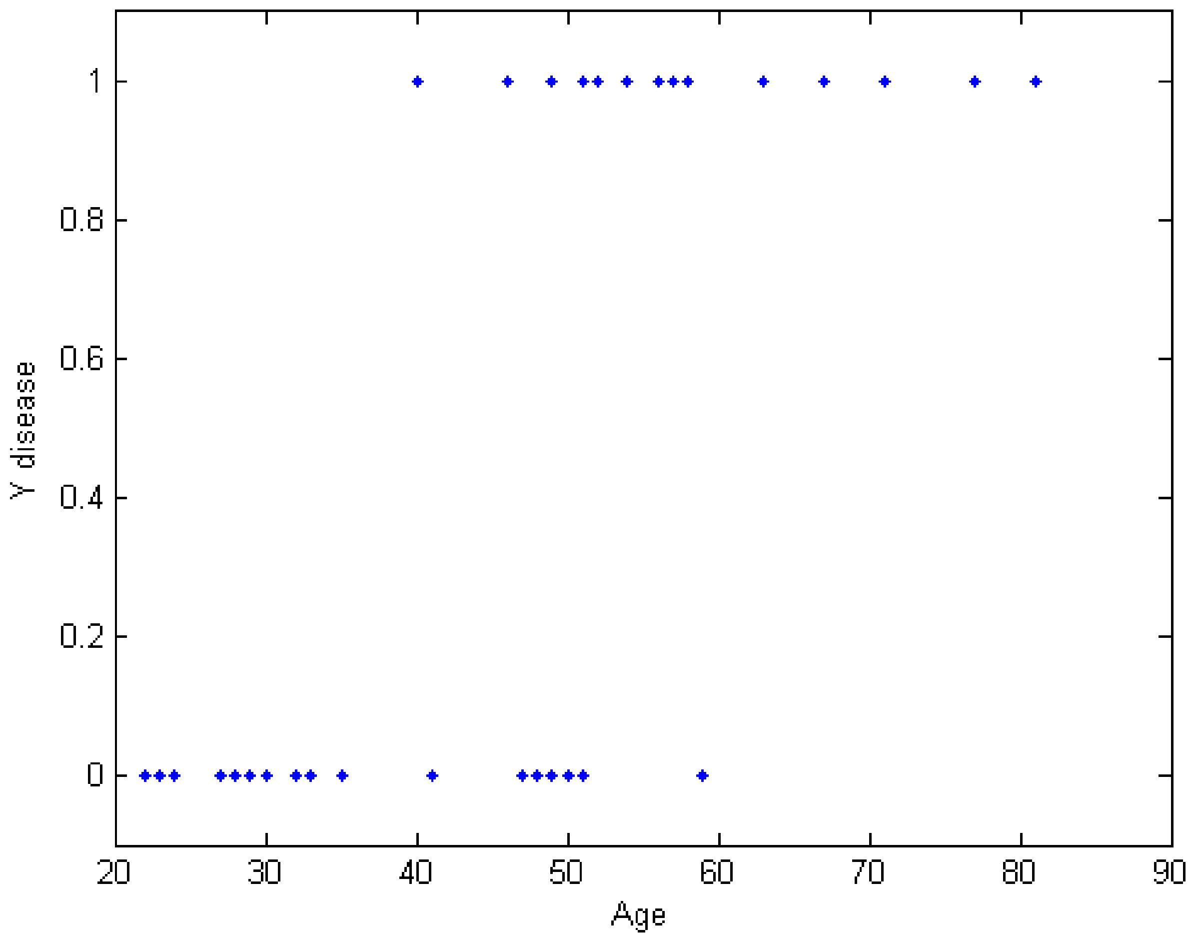
Age	Y disease
41	0
41	0
46	1
47	0
48	0
49	0
49	1
50	0
51	1
51	1
51	0

Age	Y disease
52	1
54	1
56	1
57	1
58	1
59	0
63	1
67	1
71	1
77	1
81	1

Build a model: $\text{disease} = f(\text{age}, w)$

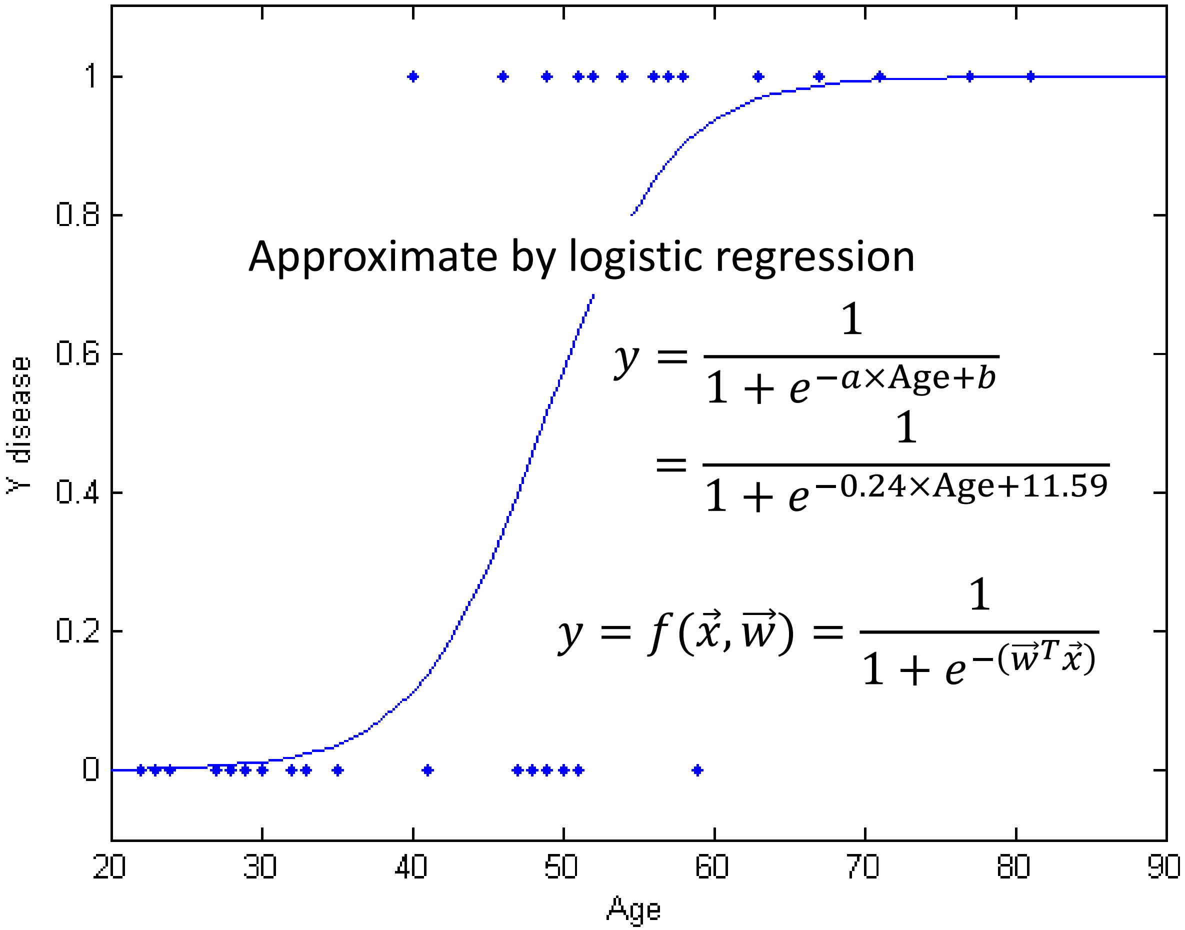
Age	Y disease
22	0
23	0
24	0
27	0
28	0
29	0
30	0
32	0
33	0
35	0
40	1

Age	Y disease
41	0
41	0
46	1
47	0
48	0
49	0
49	1
50	0
51	1
51	1
51	0



Age	Y disease
22	0
23	0
24	0
27	0
28	0
29	0
30	0
32	0
33	0
35	0
40	1

Age	Y disease
41	0
41	0
46	1
47	0
48	0
49	0
49	1
50	0
51	1
51	1
51	0



Approximate by logistic regression

$$y = \frac{1}{1 + e^{-a \times \text{Age} + b}} = \frac{1}{1 + e^{-0.24 \times \text{Age} + 11.59}}$$

$$y = f(\vec{x}, \vec{w}) = \frac{1}{1 + e^{-(\vec{w}^T \vec{x})}}$$

$$\vec{x} = \begin{bmatrix} Age \\ 1 \end{bmatrix}$$

$$\vec{w} = \begin{bmatrix} a \\ -b \end{bmatrix}$$

Logistic Regression & Classification

$$f(\vec{x},\vec{w}) = \frac{1}{1+e^{-(\vec{w}^T\vec{x})}}$$

Let $P(1|\vec{x};\vec{w}) = f(\vec{x},\vec{w})$

$$P(0|\vec{x};\vec{w}) = 1 - P(1|\vec{x};\vec{w})$$

Classification rule

$$\begin{cases} \text{Assign } \vec{x} \text{ to class 1 if } P(1|\vec{x};\vec{w}) \geq P(0|\vec{x};\vec{w}) \\ \text{Assign } \vec{x} \text{ to class 0 otherwise} \end{cases}$$

$$P(1|\vec{x};\vec{w}) = \frac{1}{1+e^{-(\vec{w}^T\vec{x})}}$$

$$P(0|\vec{x};\vec{w}) = 1 - P(1|\vec{x};\vec{w}) = \frac{e^{-(\vec{w}^T\vec{x})}}{1+e^{-(\vec{w}^T\vec{x})}}$$

$$\begin{aligned}\mathrm{logit}(\vec{x};\vec{w}) &= \log\frac{P(1|\vec{x};\vec{w})}{P(0|\vec{x};\vec{w})} \\ &= \log\frac{1}{e^{-(\vec{w}^T\vec{x})}} \\ &= \vec{w}^T\vec{x} \quad \text{Linear classification}\end{aligned}$$

Cross-Entropy Loss for Classification

A binary classification has two potential outcomes: 0 or 1. Treat binary classification as an experiment governed by a Bernoulli distribution

$$P(y_n|\vec{x}_n;\vec{w}) = \hat{y}_n{}^{y_n}(1-\hat{y}_n)^{1-y_n}$$

where y_n is the ground-truth and $\hat{y}_n = P(1|\vec{x}_n;\vec{w})$

- If $y_n = 1$, $P(y_n|\vec{x}_n;\vec{w}) = \hat{y}_n = P(1|\vec{x}_n;\vec{w})$
- If $y_n = 0$, $P(y_n|\vec{x}_n;\vec{w}) = 1-\hat{y}_n = P(0|\vec{x}_n;\vec{w})$

Cross-entropy loss: $-\log P(y_n|\vec{x};\vec{w}) = -[y_n \log \hat{y}_n + (1-y_n)\log(1-\hat{y}_n)]$

Negative *log* likelihood loss

The loss of a dataset $\{\vec{x}_n, y_n\}_{n=1\ldots N}$ is

$$-\sum_{n=1}^{N}[y_n \log \hat{y}_n + (1-y_n)\log(1-\hat{y}_n)]$$


Q
&
A