Homework 4: Logistic Regression

Introduction to Machine Learning

Shangwen Yan | N17091204 | sy2160

1.

(a)

response variable: gender of the voice(male or female)

predictors: tonal variation, speed, pause, etc.

number of classes: 2

(b)

response variable: a-z, A-Z, 0-9

predictors: how long a stylus stays on the pad without separation

number of classes: 62

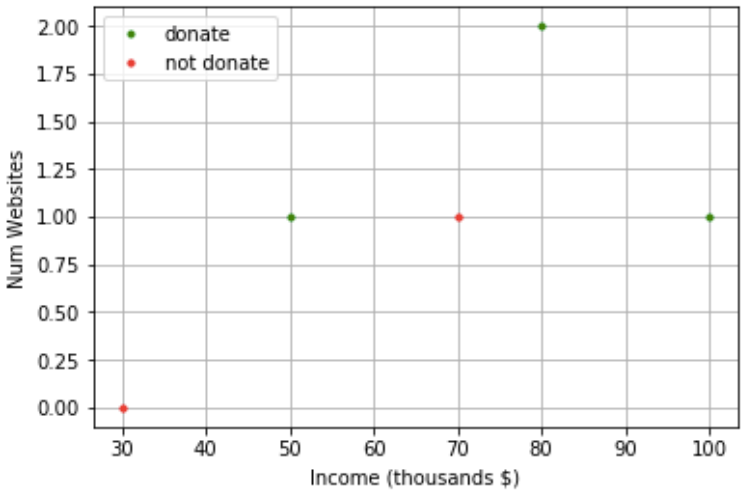
2.

(a) $\{x_1, x_2 \big| 1 + 2x_1 + 3x_2 > 0\}$

(b) $\{x_1, x_2 \big| 1 + 2x_1 + 3x_2 > ln4\}$

(c) $\{x_1 \big| x_1 > ln2 - 1.25\}$

3.



(a)

(b) $w = [-1.40], b = 29$

(c) using w and b above, first sample and the third sample are both the least likely one.

(d)

- the $\hat{y}$ in (b) will not change, cause constant c will not change the positive or nagative of z
- the likelihood in (c) will change from $\frac{1}{1+e^{z_i}}$ to $\frac{1}{1+e^{az_i}}$

(e)

```
import numpy as np
def gen_rand(X,w,b):
    y = []
    for i in range(X.shape[0]):
        if np.dot(X[i,:],w) + b > 0:
            y.append(1)
        else:
            y.append(0)
    return y
```