Introduction to Machine Learning
Unit 6 Solutions: Logistic Regression

Prof. Sundeep Rangan

1. Suggest possible response variables and predictors for the following classification problems. For each problem, indicate how many classes there are. There is no single correct answer.

   (a) Target is the gender (M or F), so there are two classes. One predictor could be the vector of digital samples of the audio. Later, we will discuss some other popular features extracted from these samples.

   (b) Target is the character written (e.g. 0-9, a-z and A-Z). So, there would be 26+26+10 classes plus classes for punctuation. The predictor could be a vector of (x, y) positions of the stylus. Each position may also include a pressure. Of course, the details would depend on the stylus interface implementation.

2. (a) Note that

$$P(y=0|\mathbf{x}) = 1 - P(y=0|\mathbf{x}) = 1 - \frac{1}{1+e^{-z}} = \frac{e^{-z}}{1+e^{-z}}.$$

   Hence,

$$P(y=1|\mathbf{x}) > P(y=0|\mathbf{x}) \Longleftrightarrow \frac{1}{1+e^{-z}} > \frac{e^{-z}}{1+e^{-z}} \Longleftrightarrow e^{-z} < 1$$
$$\Longleftrightarrow z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 > 0.$$

   (b) In this case,

$$P(y=1|\mathbf{x}) > 0.8 \Longleftrightarrow \frac{1}{1+e^{-z}} > 0.8$$
$$\Longleftrightarrow z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 > -\ln\left[\frac{0.2}{0.8}\right] = \ln(4).$$

   (c) Combining with part (b), the set is,

$$1 + 2x_1 + 3x_2 > \ln(4) \text{ and } x_2 = 0.5 \Rightarrow x_1 > \frac{\ln(4) - 2.5}{3}.$$

3. (a) The scatter plot is shown in Fig. 1.

   (b) A simple classifier is to use the boundary $x_2 = 0.5$, shown on the figure. So, we take

$$z_i = x_2 - 0.5 = \mathbf{w}^\mathsf{T}\mathbf{x} + b,$$

   where $\mathbf{w} = [0, 1]$ and $b = -0.5$.

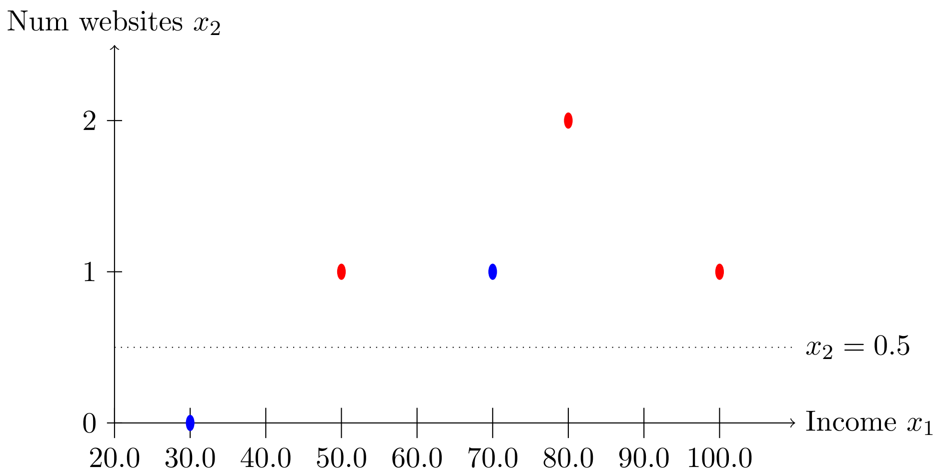

Figure 1: Scatter plot of the data points where the red circles are $y_i = 1$ and blue are $y_i = 0$

(c) We have

$$P(y_i = 1|\mathbf{x}_i) = \frac{1}{1+e^{-z_i}} \Rightarrow P(y_i = 0|\mathbf{x}_i) = 1 - \frac{1}{1+e^{-z_i}} = \frac{1}{1+e^{z_i}}.$$

Hence, we can write

$$P(y_i|\mathbf{x}_i) = \frac{1}{1+e^{-u_i}}, \quad u_i = \begin{cases} z_i & \text{if } y_i = 1, \\ -z_i & \text{if } y_i = 0 \end{cases}$$

Since $1/(1+e^{-u})$ is increasing in u, the likelihood will be minimized for the sample where u_i is the smallest. We calculate u_i for each sample using the following table:

Income (thousands \$), x_{i1}	30	50	70	80	100
Num websites, x_{i2}	0	1	1	2	1
Donate (1=yes or 0=no), y_i	0	1	0	1	1
$z_i = x_{i2} - 0.5$	-0.5	0.5	0.5	1.5	0.5
u_i	0.5	0.5	-0.5	1.5	0.5

We see u_i is smallest for sample $i = 3$, which is the misclassified point.

(d) Let z_i' be the new values of the linear discriminant under the new parameters. We have,

$$z_i' = (\mathbf{w}')^\mathsf{T}\mathbf{x}_i + b' = \alpha \left[\mathbf{w}^\mathsf{T}\mathbf{x}_i + b\right] = \alpha z_i.$$

Since $\alpha > 0$, the sign of z_i' is the same as z_i. Therefore $\hat{y}_i$ does not change. However, the probabilities do change. Since $\alpha > 1$,

$$z_i > 0 \Rightarrow z_i' > z_i$$
$$z_i < 0 \Rightarrow z_i' < z_i.$$

Hence, for samples where $P(y_i = 1|\mathbf{x}_i) > 0.5$, the probability will increase. For samples where $P(y_i = 1|\mathbf{x}_i) < 0.5$, the probability will decrease.

(e) You can use the following code:

```
import numpy as np
def gen_rand(X,w,b):
    z = X.dot(w)+b[:,None]
    p = 1/(1+np.exp(-z))
    n = X.shape[0]
    u = np.random.rand(n)
    y = (u < p)
    return y
```