1. We want to predict whether a person likes a movie. We know whether the person is male or female. From a population survey we know that 60% of people like the movie and of these 20% are males. You may assume that 50% of the population is male. Use Bayes Rule to construct a simple movie prediction classifier.

Solution

Let X=1 if the person likes the movie and 0 otherwise. Let Y=1 if they are male and 0 otherwise. By Bayes:

P(X=1|Y=y) = P(Y=y |X=1)P(X=1)/P(Y=y)

We know that P(Y=1 |X=1) = 0.2 and P(X=1)=0.6. Also P(Y=1)=0.5. So

P(X=1|Y=1) = 0.2x0.6/0.5= 0.24 and P(X=0|Y=1) = 1- P(X=1|Y=1)=0.76

P(X=1|Y=0) = 0.8x0.6/0.5=0.96 and P(X=0|Y=0) = 1- P(X=1|Y=0)=0.04

Knowing whether a person is male or female (Y=1 or Y=0) then we can predict X=0 if Y=1 since P(X=0|Y=1)> P(X=1|Y=1), and predict X=1 if Y=0 since P(X=1|Y=0)> P(X=0|Y=0).

2. Consider now a different movie prediction task. Suppose training data is available which consists of pairs (x_i,y_i), i=1,2,…,n, where y_i=1 if the ith person likes the move and 0 otherwise, and x_i=1 if the ith person is male and 0 otherwise. Describe how to construct a logistic regression classifier for this task.

Solution

A logistic classifier uses the model

P(Y=1 |X=x) = 1/(1+exp(-z))

with z=θx, θ a parameter. Given an input x, it calculates P(Y=1 |X=x) and P(Y=0 |X=x) and its prediction is Y=1 if P(Y=1 |X=x)> P(Y=0 |X=x) and Y=0 otherwise.

The classifier selects θ so as to maximize the log-likelihood of the training data. Letting d=(x_i,y_i), i=1,2,…,n denote the training data, the likelihood of this data is

P(D=d|θ) = $\prod_{i=1}^{n}\left(\frac{1}{1+\exp(-\theta x_i)}\right)^{y_i}\left(\frac{\exp(-\theta x_i)}{1+\exp(-\theta x_i)}\right)^{1-y_i}$

3. Logistic regression classification is often said to be most effective for linearly separable data. Explain what this means for a classifier with two inputs. What can happen is the data is not linearly separable ?

Solution

Let the classifier inputs be $x^{(1)}$ and $x^{(2)}$. Gather these together into vector $\vec{x}$. Logistic regression works by thresholding an intermediate variable $z(\vec{x}) = \theta^{(1)}x^{(1)} + \theta^{(2)}x^{(2)}$. When $z(\vec{x})>0$ the logistic classifier predicts an output of 1 and when $z(\vec{x})<0$ the classifier predicts 0. $\theta^{(1)}x^{(1)} + \theta^{(2)}x^{(2)} = 0$ defines a line, namely $x^{(1)} = cx^{(2)}$ with $c = -\theta^{(2)}/\theta^{(1)}$, in two dimensions. Points $\vec{x}$ for which $z(\vec{x})>0$ lie on one size of this line and points $\vec{x}$ for which $z(\vec{x})<0$ lie on the other side. 2D data is linearly separable when a line exists so that points $\vec{x}$ on one side of the line have one label and points $\vec{x}$ on the other side of the line have a different label. So when data is linearly separable, a well-designed logistic classifier should be able to predict the labels with high accuracy. When the data is not linearly separable, it may be impossible for a logistic classifier to give accurate predictions.

Note: this thinking carries over directly to higher dimensions where $z(\vec{x}) = \sum_{i=1}^{m} \theta^{(i)}x^{(i)}$ and m>2. $\sum_{i=1}^{m} \theta^{(i)}x^{(i)}$ now defines a plane that divides the space into two halves.

4. Download the file chdage.dat from:

https://www.scss.tcd.ie/doug.leith/ST3009/chdage.dat

This contains data from a medical study. Each row corresponds to a different person and consists of three columns. The first is the person ID. The second is their age. The third is whether they have heart disease or not (1 or 0). Our aim is to use age as an input/feature to predict whether a person has heart disease. Use logistic regression to construct a Matlab classifier for this data, and write a short report on your results (to include a plot of the original data overlaid with the classifier predictions and a report on the fraction of predictions which are correct).