

## CHAPTER 20

# Logistic Regression

Logistic regression is a supervised machine learning algorithm developed for learning classification problems. A classification learning problem is when the target variable is categorical. The goal of logistic regression is to map a function from the features of the dataset to the targets to predict the probability that a new example belongs to one of the target classes. Figure 20-1 is an example of a dataset with categorical targets.

The diagram shows a table with four columns. The first three columns are grouped under the label 'input variables' with a bracket above them. The fourth column is labeled 'categorical target variable' with a downward arrow pointing to it. The table has four rows. The first row is the header, and the following three rows are data entries. A bracket on the right side of the table groups the data rows under the label 'rows / records / entries'.

input variables			categorical target variable
x1	x2	x3	eye disease
4	6	3	glaucoma
2	4	1	cataract
...	...	...	...
5	8	5	none

**Figure 20-1.** Dataset with qualitative variables as output

## Why Logistic Regression?

To develop our understanding of classification with logistic regression and why linear regression is unsuitable for learning categorical outputs, let us consider a binary or two-class classification problem. The dataset illustrated in Figure 20-2 has the output  $y$  (i.e., eye disease) = {disease, no-disease} is an example of dataset with binary targets.