Overview of Logistic Regression Example

Dataset Link:

<https://www.kaggle.com/uciml/pima-indians-diabetes-database>

Dataset Content:

This dataset contains eight columns. This dataset is originally from the National Institute of Diabetes and Digestive and Kidney Diseases. The first seven columns contain various diagnostic measurements (e.g. glucose, BMI, blood pressure etc.) and the last column contains values of 0 and 1 which represent the presence or absence of diabetes.

Problem:

We would like to diagnostically predict whether or not a patient has diabetes, based on certain diagnostic measurement provided.

Solution:

We want to develop a classification model using Logistic Regression which allows us to classify if a patient suffer from diabetes or not.

Variables:

Independent variable 🡪 All features except the “Outcome” column.

Dependent variable 🡪 Presence of diabetes (1) or not (0)