Overview of Logistic Regression Example

Dataset Link:

<https://www.kaggle.com/yuqing01/breast-cancer#breast-cancer.csv>

Dataset Content:

This dataset contains 33 columns. The column “diagnosis” gives us information about the condition, with “B” for “Benign” and “M” for “Malignant” Breast Cancer. The remaining columns contain information about different features of Breast Cancer (radius of tumor, perimeter of tumor etc.).

Problem:

We are working as medical doctors at a hospital and we would like to know the probability of our patients having a benign or malignant tumor.

Solution:

We want to develop a classification model using Logistic Regression which allows us to classify with accuracy if the breast tumors of our patients are malignant and therefore we need to remove them or they are benign and we could follow more conservative methods of treatment.

Variables:

Independent variable 🡪 Various features computed from a digitized image of a fine needle aspirate (FNA) of a breast mass (radius, perimeter, concavity etc.)

Dependent variable 🡪 Benign Tumor (B) or Malignant Tumor(M)