

Assignment: Logistic regression

Learning goals

In this assignment, you:

1. learn to conduct logistic regression analysis.
2. improve your data manipulation skills in Python.

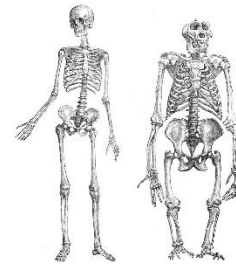


Image: public domain, via
Wikimedia Commons.

Assignment

Your task is to make a diagnostic tool (not for real medical use) that asks a medical expert six numerical quantities obtained by radiographic measurements of a patient:

1. pelvic incidence
2. pelvic tilt
3. lumbar lordosis angle
4. sacral slope
5. pelvic radius
6. grade of spondylolisthesis

As an output, your program should provide a probability estimate of the patient having a vertebral abnormality (either disk hernia or spondylolisthesis). Alternatively, your tool can not only indicate abnormality but also specify which of the two conditions an affected individual is likely to have.

Download the data set from UCI repository at <http://archive.ics.uci.edu/ml/datasets/Vertebral+Column>. There are two choices for the data set: one with a binary, and the other with a tri-categorical, response variable. Choose the one that matches your chosen objective. There's some preprocessing required: be sure to read the data description and label the variables accordingly.

In case you were interested in carrying out the actual measurements, see the Radiographic Measurement Manual at <http://www.oref.org/docs/default-source/default-document-library/sdsg-radiographic-measuremnt-manual.pdf>.

The decisions of your diagnostic tool should be based on a logistic regression model derived from the data set. If possible, provide evidence on the accuracy of your tool to convince the ever-suspicious radiological community.

Deliverables

Your deliverable should include both the Python codes and the results needed to verify the conclusions.

Submit your work preferably in pdf format.