HOCHSCHULE HEIDELBERG
Intelligence in Learning

Features of orthopedic patients

Project Report

By

Pravinkumar Konade (11010418)

2017.12.24

SRH University Heidelberg
School of Information, Media and Design

Reviewer Prof. Dr. Simon Ziegler

Table of Contents

Supervised Learning

We know classes of instances, based on that predict values. Here orthopedic patient's data that have labels of normal and abnormal.

- There are features (predictor variable) and target variable. Features are like pelvic radius or sacral slope. Target variables are labels normal and abnormal
- Given features(input) predict whether target variable(output) is normal or abnormal
- Classification: target variable consists of categories like normal or abnormal
- Regression: target variable is continuous like stock market

Exploratory Data Analysis (EDA)

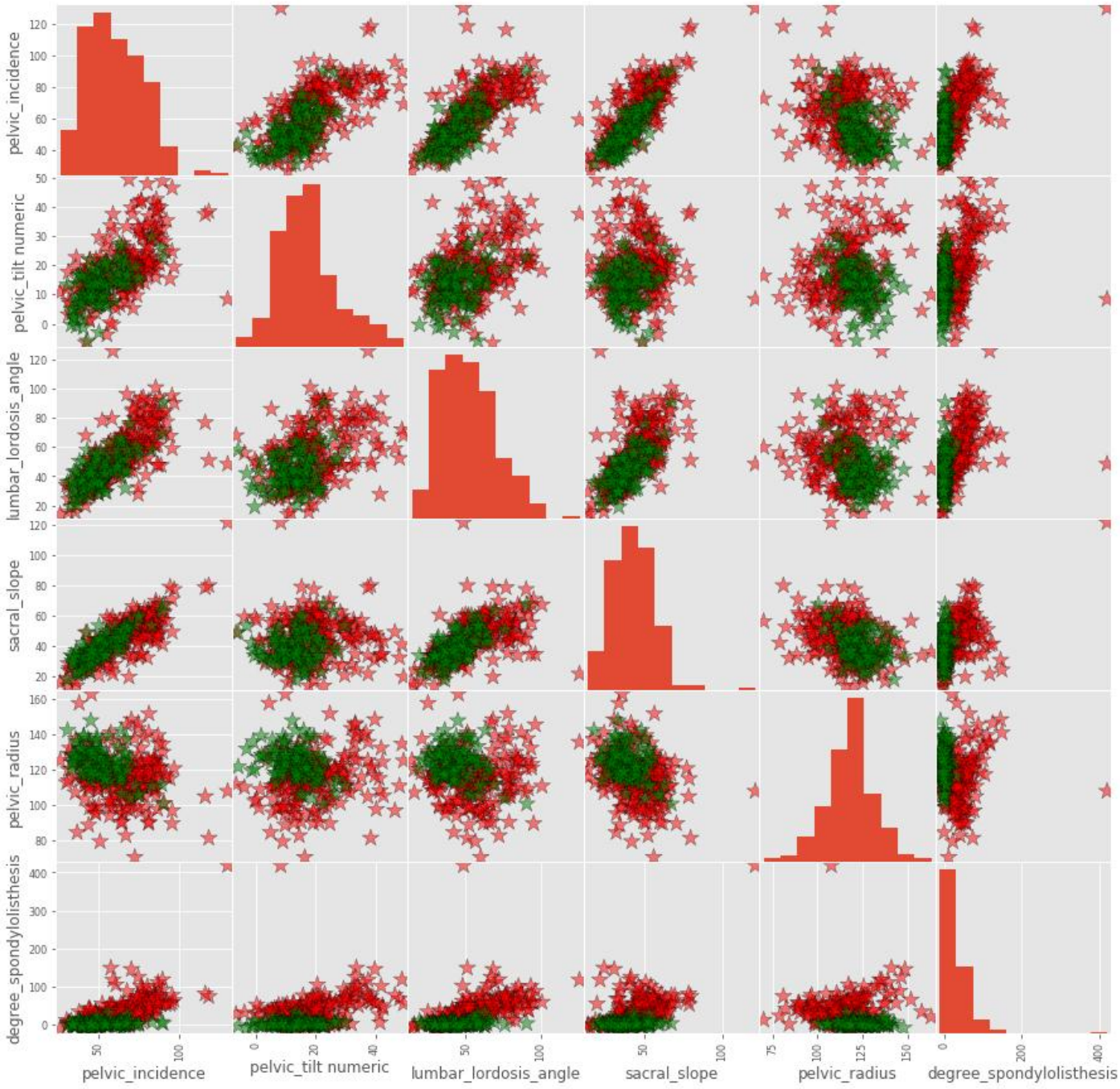
- Now we need to explore data using exploratory data analysis
- Started with head() to see features that are pelvic_incidence, pelvic_tilt numeric, lumbar_lordosis_angle, sacral_slope, pelvic_radius and degree_spondylolisthesis and target variable that is class
- head(): default value of it shows first 5 rows of samples

pelvic_incidence	pelvic_tilt numeric	lumbar_lordosis_angle	sacral_slope	pelvic_radius	degree_spondylolisthesis	Class
63.027818	22.552586	39.609117	40.475232	98.672917	-0.2544	Abnormal
39.056951	10.060991	25.015378	28.99596	114.40543	4.564259	Abnormal
68.832021	22.218482	50.092194	46.613539	105.98514	-3.530317	Abnormal
69.297008	24.652878	44.311238	44.64413	101.8685	11.211523	Abnormal
49.712859	9.652075	28.317406	40.060784	108.16873	7.918501	Abnormal

Scatter matrix:

A scatter matrix consists of several pair-wise scatter plots of variables presented in a matrix format. It can be used to determine whether the variables are correlated and whether the correlation is normal or abnormal.

- green: normal and red: abnormal
- c: color
- fig size: figure size
- diagonal: histogram of each features
- alpha: opacity
- s: size of marker
- marker: marker type


pelvic_incidence
pelvic_tilt numeric
lumbar_lordosis_angle
sacral_slope
pelvic_radius
degree_spondylolisthesis

K-Nearest Neighbors (KNN)

Feature vectors of the training set are called “samples”. A feature vector of the testing set is classified by just determining the nearest sample. The class of the sample is the classification of the feature vector which should be classified. Because there is no training algorithm, NN is called “lazy learning”. Not all samples are necessary for the K-NN classification. Just samples near class-borders are important

- KNN: Identify K closest labeled data points
- Classification method.
- First we need to train our data. Train = fit
- fit() : fits the data, train the data.
- predict(): predicts the data
- x: features
- y: target variables(normal, abnormal)
- n neighbor’s: finding K it means that Look at the 3 closest labeled data points

Principle Component Analysis (PCA)

The principle component analysis (PCA) is a method for describing main axes in datasets of arbitrary dimensions. The main axis’s point in directions of the distribution of the datasets.

- first step is deco-relation:
- rotates data samples to be aligned with axes
- shifts data samples so they have mean zero
- no information lost
- fit() : try to shift samples
- transform(): apply the learned transformation. It can also be applies test data
- Resulting PCA features are not linearly correlated
- Principle components: directions of variance

Executive Summary

This project is about the implementation of such Features of orthopedic patients. The task consists in classifying patients as belonging to one out of two categories: Normal or Abnormal. Implementing supervised learning, K-nearest neighbor algorithm, Feature Computation, Exploratory data analysis etc.

Bibliography

1) https://archive.ics.uci.edu/ml/index.php
2) https://www.youtube.com/watch?v=K_oXb04izZM
3) https://www.coursera.org/learn/machine-learning
4) https://www.wikipedia.org/