CS 777 Final Project

Biomechanical Features of Orthopedic Patients

Nathan Oasis

May 2020

Contents

1 Introduction

Physical health is incredibly important. By examining biomedical features through ML, we can quicken the diagnosis process and better connect patients to the professional help they need.

2 Data Description

We have obtained a biomedical dataset hosted by the University of California. In this section, we shall provide a brief description of the classes, their distributions, the distributions of the features, and a resource to get more information on the features. [2]

2.1 Feature Set Description

[1] presents a detailed description of each of the features. Succinctly, each feature is a length or angle from one part of the body to another.

2.2 Feature Set Metadata

See Table 1. Each feature has been normalized by class.

2.3 Class Description

There are three unique classes: Normal, Hernia, and Spondylolisthesis. Additionally, the Hernia and Spondylolisthesis classes can be considered as one class, named, Abnormal. A Hernia is refers to the protrusion of an organ through the wall of the cavity containing it. Spondylolisthesis refers to the slippage of one vertebra over another. [1]

2.4 Class Distribution

See Table 2.

summary	pelvic_incidence	pelvic_tilt	lumbar_lordosis_angle	sacral_slope	pelvic_radius	degree_spondylolisthesis
count	310	310	310	310	310	310
mean	60.4966	17.5428	51.9309	42.9538	117.9206	26.2967
stddev	17.2365	10.0083	18.5540	13.4231	13.31737	37.5590
min	26.1479	-6.5549	14.0	13.3669	70.08257	-11.0581
max	129.8340	49.4318	125.74238	121.4295	163.07104	418.5430

Table 1: Metadata on the features of the dataset.

Class		Count
Normal		100
Abnormal	**Hernia**	60
	Spondylolisthesis	150
		310

Table 2: Class Distribution of Data.

3 Graphical Exploration

Before designing the project, we perform a preliminary graphical exploration of the features with hopes to gleam novel insights into the data and thus how to best direct our efforts. From examining the pairplot, included in the accompanying jupyter notebook, we see the **degree_spondylolisthesis** appears to have significant discriminatory power. Further, the data appears linearly separable. Work done in [1] also suggests the data is linearly separable.

4 Project Design

4.1 Research Questions

This research attempts to answer the following questions:

1. Which of the features discussed in Section 2.1 can be used to accurately classify the **normal** back type and the **abnormal** back type with results statistically significant to our baseline?

2. Which of the features discussed in Section 2.1 can be used to accurately classify the **abnormal** back types with results statistically significant to our baseline?

3. Which of the features discussed in Section 2.1 can be used to accurately classify all back types with results statistically significant to our baseline?

4.2 Dataset Construction

See table 3.

4.3 Experiment Design

From our observations in Section 3 we will use linear classifiers, specifically Spark's `LogisticRegression` to answer our questions as posed. Questions 1 and 2 involve two classes each, while question 3 involves three classes. Thus, we will be using a binary classifier and multinomial classifier, respectively, to learn our data and answer the questions.

	Dataset One	Dataset Two	Dataset Three
Normal	100		100
Hernia		60	60
Spondylolisthesis		150	150
Abnormal	210		
Total	310	210	310

Table 3: The class distribution for each dataset. Each dataset corresponds to a research question. Dataset One separates all the data into two classes: Normal and Abnormal. The Abnormal class is the union of the Hernia and Spondylolisthesis classes. Dataset Two contains only the Abnormal class subsets of Hernia and Spondylolisthesis. Dataset Three contains all class labels Normal, Hernia, and Spondylolisthesis.

5 Results

Table 4 details our experiments on each dataset. We have demonstrated these features encapsulate non-trivial information predictive of the labels.

		Algorithm	
Dataset	Instance Count	LogRegression	ZeroR (Baseline)
Dataset One	310	**0.8434**	0.6774
Dataset Two	210	**0.9760**	0.7143
Dataset Three	310	**0.8283**	0.4839

Table 4: The FMeasure score for each dataset using logistic regression compared to a ZeroR baseline.

5.1 Answering Research Questions

To understand which features best distinguish the classes, we examine the coefficients after the model has finished training. Table 5 shows the results for `DATASET_THREE`. These do contain most of the results for the first two datasets, but we encourage the reader to follow the code in the accompanying jupyter notebook.

Examining the coefficients created from performing logistic regression on `DATASET_ONE` revealed the pelvic_radius and the degree_spondylolisthesis were the primary influences of differentiating between the **Normal** and **Abnormal** classes.

When examining the coefficients created from performing logistic regression on `DATASET_TWO`, it was revealed the sacral_slope contained the most discriminatory power. The sacral_slope which was hardly a factor in the first dataset sits at nearly double the weight of degree_spondylolisthesis in this classification task. The sacral_slope, and the degree_spondylolisthesis were the primary influences of differentiating between the **Hernia** and **Spondylolisthesis** classes.

The coefficients created from performing logistic regression on `DATASET_THREE` (Refer to Table 5.) retold some of the information we had learned from our first two experiments. Similar to experiment one, degree_spondylolisthesis is the primary differentiator between the **Normal** and **Spondylolisthesis** classes, though interestingly the feature has no bearing on the **Hernia** class. Like in experiment two, the sacral_slope contained the most discriminatory power for the **Hernia** class.

class	pelvic_in	pelvic_tilt	lumbar_ang	sacral_m	pelvic_r	degree_s
N	0.	-0.0148	0.	0.	0.0087	-0.0122
H	0.	0.0311	-0.0192	-0.0719	-0.0052	0.
S	0.	0.	0.0106	0.0151	-0.0168	0.0784

Table 5: The multinomial coefficients of the features by class. This table depicts the weight a feature has for predicting a given class.

References

[1] Eric Berthonnaud et al. "Analysis of the Sagittal Balance of the Spine and Pelvis Using Shape and Orientation Parameters". In: *Journal of spinal disorders & techniques* 18 (Mar. 2005), pp. 40–7. DOI: `10.1097/01.bsd.0000117542.88865.77`.

[2] Dheeru Dua and Casey Graff. *UCI Machine Learning Repository.* 2017. URL: `http://archive.ics.uci.edu/ml`.