Classifying Disk Herniation Diseases

Maya Maya Barbosa Silva (up201700077)
Jacinta Dias Ferreira (up201705451)
Simeon Roshchupkin (up201902658)

1.Introduction

Disk herniation is the extrusion of the core of the intervertebral disk. It causes back pain and can even affect nerve activity. Diagnosis can be done by test as x-ray and magnetic resonance imaging. [2]

Spondylolisthesis is the dislocation of a vertebra in relation to another one, likely caused by stress or fatigue. Its most common location is the L5 vertebra. It is the most common identifiable cause of back pain in children, and it frequently affects adolescent athletes.[1] The diagnosis is usually done by imaging tests.

There are several attributes that can be evaluated to determine if a patient has one of these conditions. The pelvic incidence is the angle between a line drawn from the center of the femoral head to the midpoint of the sacral endplate and a line perpendicular to the center of the sacral endplate. The pelvic tilt is the angle between the vertical reference line and the line from the middle of the sacral endplate to the axis of the femoral heads. The lordosis angle is the bigger sagittal angle between the sacrum superior plate and the lumbar vertebra superior plate. The sacral slope is the angle between the sacral endplate and the horizontal reference line. The pelvic radius is drawn from the center of the line connecting the centers of the femoral heads to the center of the sacral endplate. Finally, the grade of spondylolisthesis is the percentage of slipping between the inferior plate of the fifth lumbar vertebra and the sacrum.

Machine learning can be used to aid the diagnosis of these conditions, improving the efficiency of the diagnosis process.

Naïve Bayes is a classifier based on Bayes' Theorem, in which the features are assumed to be independent. Bayes' theorem states that

$$P(Ck|x) = P(Ck)P(x|Ck)P(x)$$

The Mixture of Gaussian process used assumes that the data follows a normal distribution and that it can be generated through a combination of several gaussians with variable parameters.

K-Nearest Neighbor is a non parametric classifier in which the objects are assigned to the most common class among their k nearest neighbors. The value of k impacts the classification.

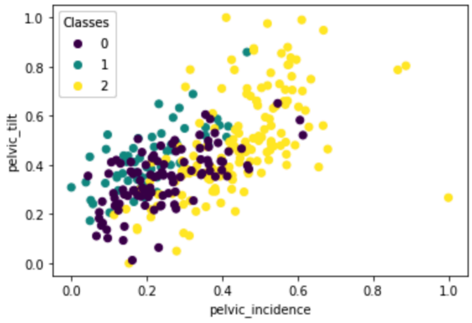
2.Methodology

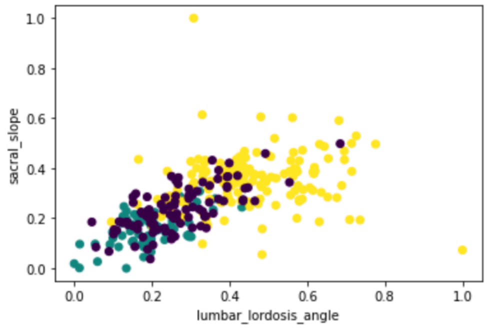
The database used for this work provides values for these six attributes. The classification is done in two ways. On the first alternative the patients are classified as “normal” or “abnormal”. On the second there are three categories: “normal”, “disk hernia” and “spondylolisthesis”. The database contains data for 100 normal patients, 60 patients with disk hernia, and 150 patients with spondylolisthesis. Three algorithms were used to perform the classification of the data: Naive Bayes, Gaussian Process and K-Nearest Neighbor.

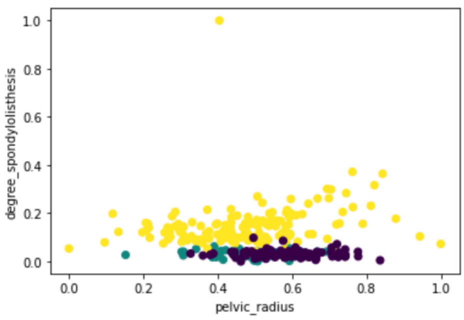
For the first part of the dataset, the classes, “normal” and “abnormal” are labeled as “0” and “1”, respectively. For the second part of the dataset, the classes, “normal”, “hernia” and “spondylolisthesis”, are labeled as “0”, “1”, and ”2”. 80% of the data is randomly separated to be used in training, and the other 20% is left for testing.

2.1. Naive Bayes

For classifying objects with non-binary features, the most suitable of Naive Bayes implementations is a Gaussian Naive Bayes(GaussianNB) algorithm presented in scikit-learn library. Initial dataset splitted to parts for training and testing, and features data normalized by min-max way to reduce possibility of incorrect features weighting in trained model (we have 6 features with different range of values). All 6 features can be graphically represented by three 2-d graphs:





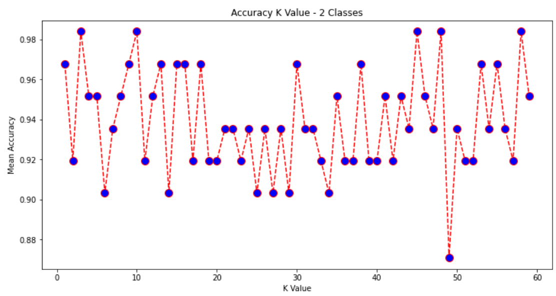


2.2. Gaussian Process

The pre-existent scikit-learn python library *GaussianMixture (GMM)* was used to generate a Gaussian Mixture algorithm respecting the 80% for training and 20% for testing convention . Each class had its mean calculated and then used to initialize the GMM process. Finally, the results were obtained by comparing the predicted values given by the GMM formula to the original classification.
The same method was used for both parts of the dataset.

2.3. K-Nearest Neighbor

The scikit-learn library is used to train a k-nearest neighbor algorithm and to apply it to the test data. To choose the number of neighbors, the algorithm was used with k equal to all numbers from 1 to 40, and the accuracy of the test was calculated, by comparing the labels contained in the dataset to the labels obtained by the algorithm. To prevent bias caused by the randomization of the data, this process is repeated 50 times, with data randomly selected each time, and the average accuracy is calculated. The result can be seen in figure X.



Since there seems to be no optimal value, the value of k was calculated by the square root of the number of the number of samples

$$k = \sqrt{61} = 7,81$$

The number of neighbors is, then, set to 8. The algorithm is used again and the accuracy, sensitivity and specificity are displayed.

3.Results and Discussion

3.1 Naive Bayes Classification:

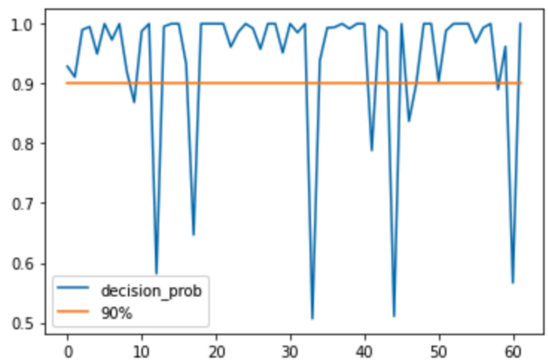
The main problem that affects the process of training and evaluating model, is a small dataset, the only 310 rows doesn't allow us to separate many samples for appropriate testing, because all of them are valuable for training, while using a very small set for testing a model isn't representative. To reduce this factor training was repeated a number of times with the different randomly splitting dataset and on each iteration accuracy (misclassification error) were calculated:

#	random_state	Accuracy (test_set)	Accuracy (train_set)	Sensitivity	Specificity
1	1	0.822581	0.778226	0.79	0.90
2	5	0.774194	0.782258	0.70	0.91
3	42	0.790323	0.782258	0.77	0.83
4	66	0.709677	0.798387	0.60	0.91
5	123	0.741935	0.794355	0.69	0.85
mean		0.767742	0.787097	0.71	0.88
st.dev		0.043578	0.008742	0.07	0.04

The same approach were applied for the second dataset where three classes of labels are selected:

#	Accuarcy (test_set)	Accuarcy (train_set)	presicion(test)			recall(test)		
			0	1	2	0	1	2
mean	0.819355	0.841935	0.77	0.66	0.94	0.71	0.71	0.95
st.dev	0.034967	0.007756	0.07	0.04	0.05	0.07	0.10	0.04

We also can observe what probability was gained in the model for every decision and as we can notice, many decisions are made by model with probability less than 90% (for example, in one of the experiments we can observe that some probabilities are even less 60%).



3.2 Gaussian Process

Part I: “Normal” or “Abnormal”

	Statistics	
	Mean	Std.Dev
Accuracy Train (%)	51,693548	17,0710778
Accuracy Test (%)	51,612903	17,2059811
Specificity	0,0684717	0,05569656
Sensitivity	0,9347842	0,03679685

To the first part of the project, several tests were done in order to obtain a mean value for the accuracy. As observed in Table 1, the results varied significantly from each test to another. This phenomenon was interpreted by the students as being the influence of the data shuffling, creating some situations in which the train and test set had enough heterogeneity to produce a better classification as well as others in which the sets could contain a high percentage of one of the classes in comparison to the others, setting a “bias” to the Classifier.

Part II: “Normal”; “Hernia” or

	Statistics		
	General		
		Mean	Stand. Dev
Accuracy Train (%)		68,1451613	17,9718425
Accuracy Test (%)		68,7096774	19,7407758
	Class:		
	Class 0	Class 1	Class 2
Specificity	0,97916667	0,8974359	0,9373761
Sensitivity	0,484375	1	0,01818182
Class Accuracy	0,9516129	0,90322581	0,70967742

“Spondylolisthesis”

Similarly to the first part, a Multi-Class Confusion Matrix was created in order to allow the calculus of the specificity, sensitivity and Accuracy of the method. The overall results of this part were better than the previous one. This conclusion was interpreted as being the consequence of less data overlay, resulting in an easier task to classify.

3.3 k-Nearest Neighbor

Part I: “Normal” or “Abnormal”

The algorithm was run 8 times, and the results for accuracy, specificity and sensitivity are displayed on Table 3.

	accuracy	specificity	sensitivity
mean	0,9415	0,5313	0,9465
std. dev	0,0258	0,5078	0,0147

The values of accuracy and sensitivity were both very good. However, the specificity had a very low average and high standard deviation, meaning the values vary a lot.

Part II: “Normal”; “Hernia” or “Spondylolisthesis”

The results for the 8 tests can be seen on Table 4.

		accuracy			specificity			sensitivity		
	accuracy	0	1	2	0	1	2	0	1	2
mean	0,92	0,95	1,0	0,94	0,25	0,89	1,00	0,95	0,94	0,96
std. dev	0,04	0,03	0,0	0,02	0,46	0,06	0,02	0,03	0,04	0,01

As in the first part, the value of general accuracy, and the values of accuracy and sensitivity for the different classes were very good. However, the specificity for class 0 varies greatly and sometimes reaches very low values.

In general, this algorithm has a good performance in the classification of the data. The only considerable limitation seems to be related to the objects belonging to class 0, which are often misclassified.

3.4 Algorithm comparison

By comparing the results for Gaussian Process and K-Nearest Neighbor, the group could conclude that the K-Nearest Neighbor accuracy values of the classification are better and, generally, more consistent. The same can be said to the sensitivity. However, it is important to consider that both algorithms produce Specificity Values with a high Standard Deviation.

Conclusion

Application machine learning algorithms in solving biomedical problems give us an almost limitless field for experiments in the use of a variety of algorithms, its compositions, and the ability to adjust parameters. But as we can notice during the process of developing a project, one of the key factors is the volume of data that can be used for training, testing, and validating. In our experiments, we received a wide range of values of accuracies and errors, as well as model probabilities for decisions sometimes not big enough, which points that trained models aren’t very “confident” in making decisions. Partially this problem can be reduced by applying methods of the artificial increasing size of datasets.

In addition, it is important to note that unlike many other areas of application of computer algorithms, the diagnosis of patients requires not only high accuracy, but also the ability to demonstrate decision-making criteria. Such a class of problems is considered as explainable artificial intelligence and this important requirement should always be taken into account in developing computer-aided diagnostics systems.

5.References

[1] Foreman, P., Griessenauer, C. J., Watanabe, K., Conklin, M., Shoja, M. M., Rozzelle, C. J., … Tubbs, R. S. (2012). L5 spondylolysis/spondylolisthesis: a comprehensive review with an anatomic focus. Child’s Nervous System, 29(2), 209–216. doi:10.1007/s00381-012-1942-2

[2] Oyedotun, O. K., Olaniyi, E. O., & Khashman, A. (2016). Disk hernia and spondylolisthesis diagnosis using biomechanical features and neural network. Technology and Health Care, 24(2), 267–279. doi:10.3233/thc-151126