Probabilistic Methods for Diagnosing Parkinson's Disease in Hand-Drawn Spirals

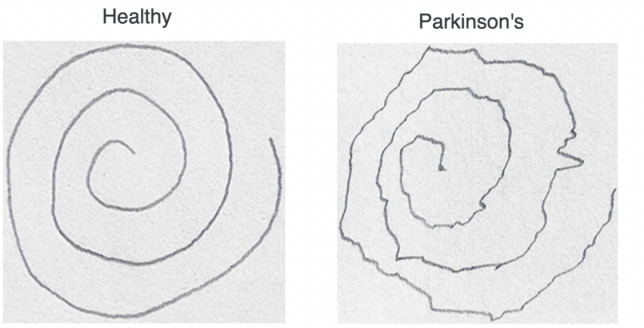
Rohan Sikand

CS 109, Winter 2022

Video link.

1 Introduction

We propose several different methods for identifying Parkinson's disease through drawings of spirals. In essence, this is a classification task: given an image of a hand-drawn spiral, does the patient, who drew the image, have Parkinson's disease?



Parkinson's disease is a disease of the nervous system that affects movement. It often leads to instability in movements and tremors. The diagnosis of Parkinson's disease is ambiguous: there is no specific test. A trained medical doctor makes a diagnosis through a historical analysis of patient symptoms, tests, and other factors. The problem is that a trained medical professional isn't always available–especially in under-resourced, remote areas. Thus, there is a need for an efficient heuristic for the diagnosis of Parkinson's disease. here we propose to study hand-drawn spirals using various concepts from probability theory.

2 Methods

We developed and tested several methods for this classification task–all based on probability theory.

2.1 Data

The data was collected from Kaggle (link) and the structure of the data is as follows:

- Split into `train` and test.
- For both 'train' and 'test', there are two classes of images: `healthy` and `parkinson`.
- There are 36 training samples for each class and 15 test samples for each class.

2.2 Probabilistic modeling of curvature

Here we propose an interesting paradigm: using **curvature as a measure**. Eventually, we will create a distribution from curvature measures but let's first discuss how to measure curvature.

The motivation is that the "curvature" of the drawing from a Parkinson's patient might be a bit more jagged than that from a healthy patient. We will use probability to analyze such curvature but we first need a way to actually measure the curvature of the drawing from the image.

For each image, a segmentation process, via thresholding (by determining which pixel's had grayscale values that were below the value 230) was performed to identify which pixel values had pencil present (i.e. part of the drawing). This yielded a collection of Cartesian (x, y) coordinates that compose a curve that resembles the hand drawing from the image.

Given these (x, y) coordinates for each sample, we can calculate the "curvature" for each curve by using ideas from Calculus (see here and here for more details on this calculation).

Now that we have a measure of curvature for every sample, it is time to use probability theory to predict classifications.

Let X and Y be random variables that represent the measure of curvature for healthy patients and parkinson's patients respectively. From the training data, we approximate these random variables as normals (using unbiased estimations for the parameters). For each sample in the training data, a curvature was measured and appended to one of two new n vectors which we denote as $\mathbf{v} = [v_1, ..., v_n]$ and $\mathbf{w} = [w_1, ..., w_n]$ for healthy patients and parkinson's patients respectively. Thus,

$$X \sim N\left(\mu = \frac{1}{n}\sum_{i=1}^{n} v_i, \sigma^2 = \frac{\sum (v_i - \bar{v})^2}{n-1}\right), Y \sim N\left(\mu = \frac{1}{n}\sum_{i=1}^{n} w_i, \sigma^2 = \frac{\sum (w_i - \bar{w})^2}{n-1}\right)$$

Now that we have our random variable distributions approximated as normals from the training data, we can make predictions on the test data. Our approach for this was as follows: (1) for each test sample, get the curvature measure, denote it as c_i; (2) calculate the probability density from the probability density function for each $X = c_i$ and $Y = c_i$; (3) divide the two probability densities to get a ratio and use that ratio to determine the prediction. We derive this mathematically as follows (in summary, we are interested in finding the ratio of probability densities[1]):

$$\frac{f_x(c_i)}{f_y(c_i)}.$$

From here, we check if $\frac{f_x(C_i)}{f_y(C_i)} \geq 1$. If so, the probability density ratio was above 1:1 and so the class prediction for the sample c_i is the class represented by X: `healthy`. If the ratio was below 1, then the prediction is the class represented by Y: `Parkinson`.

Utilizing this approach on the test data, we predicted 18/30 samples correctly for a total accuracy of 60%. Not terrible. But we can do better.

2.3 Logistic regression

We perform logistic regression on the images directly as samples. Some pre-processing was performed on the data: specifically, each pixel value was made binary (no pencil, 0, or pencil present, 1) based on a threshold value (if the grayscale value of the current pixel was below 230, it was set to 1 and and if greater than 230, it was set to 0). Second, all of the image samples for the training and testing sets were downsampled to reduce the size of the image by a factor of 16. Each image was originally 256×256 and after downsampling, 16×16 for a total of 256 features per input[2].

Logistic regression is essentially a function approximation that maps $\mathbf{x}$ to y. This is the same thing as $g(\mathbf{x}) = \text{argmax}\, \hat{P}(Y = y \mid \mathbf{X})$. Mathematically, each prediction is made based on the trained values for the parameters (θ's) as follows:

$$P(Y = 1 \mid \mathbf{X} = \mathbf{x}) = \sigma\left(\sum_i \theta_i x_i\right)$$

In this case, we denote the event $Y = 1$ as a prediction of the `parkinson` class and $Y = 0$ as a prediction of the `healthy` class where $\mathbf{X}$ is a list of the feature vectors (inputs). This is a binary

[1] This part of the approach is similar to problem set 4, question 7.

[2] The reasoning for doing this is because the number of features per input would have been to large to train locally.

classification problem and so we have:

$$P(Y = 1 \mid \mathbf{X} = \mathbf{x}) = \sigma\left(\theta^T \mathbf{x}\right),$$
$$P(Y = 0 \mid \mathbf{X} = \mathbf{x}) = 1 - \sigma\left(\theta^T \mathbf{x}\right)$$

This defines the forward pass which is determined based on the values for θ, the model's parameters. We learn these parameters by training the algorithm on the training data. To do this, we use the gradient ascent algorithm by updating the weights (θ's) at iteration.

To make things flow mathematically[3], we interpret a class prediction as an indicator random variable $Y \sim \text{Bern}(p)$ where $p = \sigma\left(\theta^T \mathbf{x}\right)$. We need a way to measure the performance of how well we are estimating these parameters. For that, we use the likelihood of the continuous version of the probability density function for a Bernoulli defined as

$$L(\theta) = \prod_{i=1}^{n} \sigma\left(\theta^T \mathbf{x}^{(i)}\right)^{y^{(i)}} \cdot \left[1 - \sigma\left(\theta^T \mathbf{x}^{(i)}\right)\right]^{\left(1-y^{(i)}\right)}$$

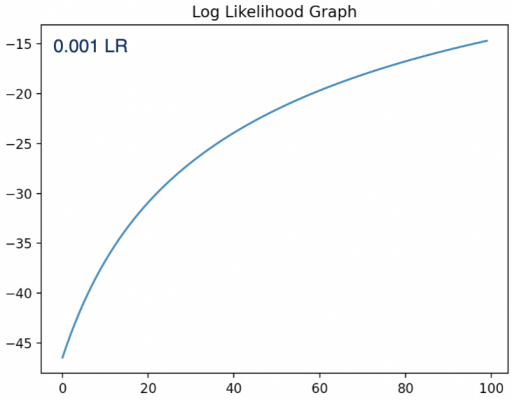
For numerical stability, we convert this to log space and define the log-likelihood function as follows:

$$LL(\theta) = \sum_{i=1}^{n} y^{(i)} \log \sigma\left(\theta^T \mathbf{x}^{(i)}\right) + \left(1 - y^{(i)}\right) \log\left[1 - \sigma\left(\theta^T \mathbf{x}^{(i)}\right)\right]$$

Now to train the algorithm using gradient ascent, we iterate over the data a number of times and at each point, update our parameters for the model. Specifically,

$$\theta_j^{\text{new}} = \theta_j^{\text{old}} + \eta \cdot \frac{\partial LL\left(\theta^{\text{old}}\right)}{\partial \theta_j^{\text{old}}} = \theta_j^{\text{old}} + \eta \cdot \sum_{i=1}^{n} \left[y^{(i)} - \sigma\left(\theta^T \mathbf{x}^{(i)}\right)\right] x_j^{(i)}.$$

We trained the model for 100 steps (epochs) with a learning rate $\eta = 0.001$. These hyperparameters were determined through experimentation. We plot the log-likelihood for the model predicting on the training data at each step in the graph below:



We can see that the model's performance was increasing throughout the entirety of training and was slowly converging.

After the model was trained by following the above approach, we predicted on the test samples. The logistic regression approach was able to achieve an **accuracy score of** 86.7%! Quite the improvement!

[3]The following logistic regression derivations were mainly adapted from the CS 109 Course Reader.

3 Conclusion

We have shown here that it is possible to achieve a relatively high accuracy in "diagnosing" (as a heuristic) patients with Parkinson's by using different probabilistic methods. We started with the assumption that the curvature of the spirals can be approximated with normal random variables. Aiming to improve performance, we then applied a different approach using logistic regression on the images themselves which resulted in an overall increase in accuracy.

Future work would involve expanding these methods (particularly the method described in section 2.2) to work for other datasets with a limited number of samples (as we had here). We believe that utilizing our own intuition (e.g. recognizing the difference in curvature between spirals), combined with an informed probabilistic framework is a step forward in making machine learning work with less data.