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

CS 109 Challenge

Rohan Sikand

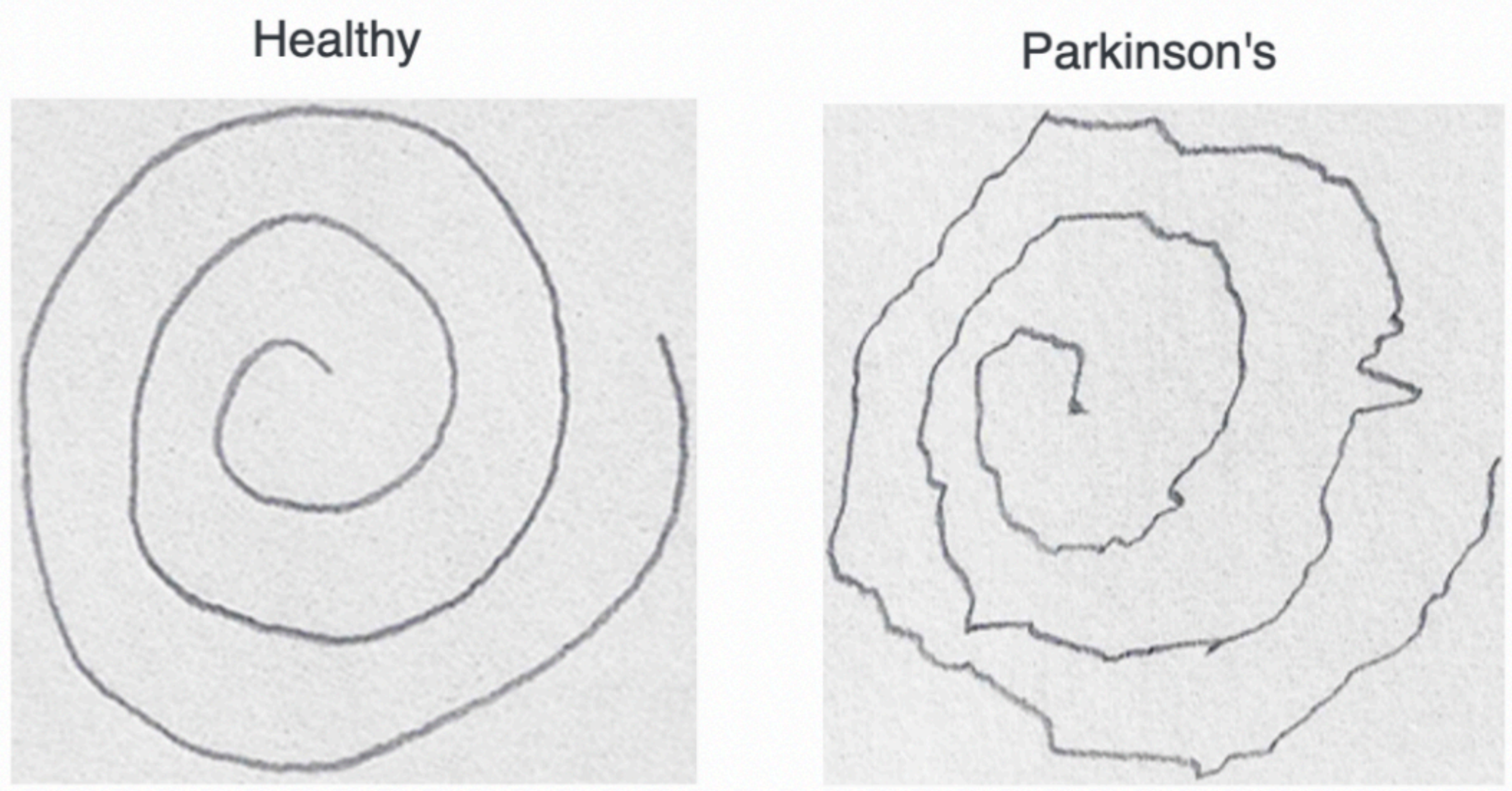
The problem

- Parkinson’s is a motor disease of the central nervous system that affects the patient’s movements and motions. Things such as tremors often occur.

- Diagnosing Parkinson’s disease is a difficult, subjective task that requires a trained medical professional. However, a trained medical professional isn’t always available—especially in under-resourced, remote areas. So here, we propose a heuristic for “diagnosing” Parkinson’s disease by probabilistically analyzing hand-drawn spirals.

The task

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?



Method 1: Probabilistic modeling of curvature

Curvature calculation:

1. Speed $= \frac{ds}{dt} = |\mathbf{v}(t)| = \sqrt{(x')^2 + (y')^2}$.
2. $\mathbf{v} = \frac{ds}{dt}\mathbf{T}, \quad \mathbf{T} = \frac{\mathbf{v}}{ds/dt}$
3. $\mathbf{a}(t) = \frac{d^2 s}{dt^2}\mathbf{T} + \kappa\left(\frac{ds}{dt}\right)^2\mathbf{N} = \frac{d^2 s}{dt^2}\mathbf{T} + \frac{v^2}{R}\mathbf{N}$
4. $\kappa = \frac{d\mathbf{T}}{ds} = \frac{|\mathbf{a}\times\mathbf{v}|}{|\mathbf{v}|^3}$. 4a. For plane curves $r(t) = x(t)\hat{\mathbf{i}} + y(t)\hat{\mathbf{j}}$: $\quad \kappa = \frac{|x''y' - x'y''|}{\left((x')^2 + (y')^2\right)^{3/2}}$.

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

$$\frac{f_x(C_i)}{f_y(C_i)} \geq 1.$$

Healthy distribution	Parkinson's distribution
2.7711527622076604	2.541980494838526
3.4737810516049783	3.133604209795485
4.037943698449856	3.2079562723166117
2.5182261850189045	3.3200364892348033
3.340271126714749	3.994433123264439
3.791156312661872	3.9034384122252854
3.2250570434994	3.1808788295647537
3.5609485450615406	3.818119750873864
3.610568362483777	3.167016993387118
3.597382932680112	4.1802575053677415
3.6570681454680907	3.2864347662455176
3.1096810423075696	3.2702385991480036
3.0060156963177342	2.6547941025582253
3.8517123027144065	3.5225961233671494
3.954610195829943	3.5647051506449774
3.2663074859642056	2.999859367382493
3.8337460798978045	3.3194101939887624
3.3631454169390578	3.12588349303676
3.9916827479678014	3.5588810114395115
3.5119212452811537	3.244064800971464
2.847211097768212	3.276143377349889
2.7056251398939573	2.66242628817899
2.450988228353994	2.4857039119365147
2.600622874288327	2.0845185055984943
3.9171950834682456	3.8251139147208333
3.259022470015326	2.6526872477426036
3.3632718646280995	4.008111135641105
3.228996391538982	3.6261646152512217
2.7180583181993336	2.9852472385019557
3.2223652565479797	3.0957511734501257
3.262915156762849	3.2443868465756625
3.48400055501952	3.9706772780157977
2.9074306812813506	3.2921141705309527
2.0949307843013028	2.6987976188043445
3.4486977772615632	3.304590802929914
3.0451835797249647	3.695041428900466

Demo 1

Method 2: Logistic Regression

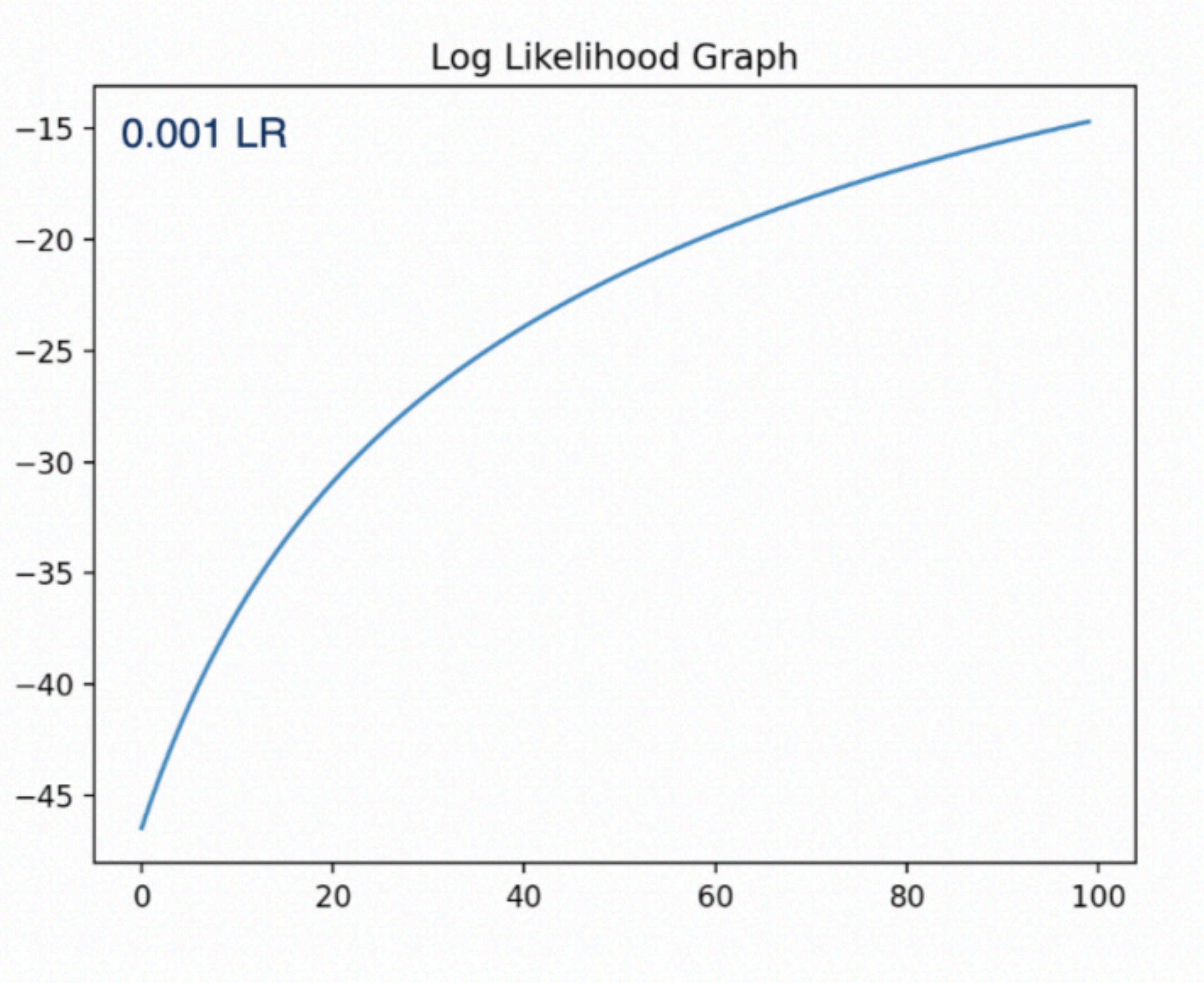
Y = 1 as a prediction of the Parkinson class*

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

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

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

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



Demo 2

Results + Conclusion

- Method 1 (curvature as a measure): 60% accuracy
- Method 2 (logistic regression): 86.7% accuracy
- We have shown that it is possible to achieve high accuracy on small amount of data (< 100 training samples) by using probabilistic methods
- Useful heuristic for diagnosing Parkinson's—helpful in under-resourced areas.
- Future: few-shot learning by understanding and incorporating the semantics of the question being asked?