Report 2: Gaussian Process Regression on Parkinson's disease data

Marco Colocrese, s301227,
ICT for Health attended in A.Y. 2021/22

November 25th, 2021

1 Introduction

Patients affected by Parkinson's disease cannot perfectly control their muscles. In particular they show tremor, they walk with difficulties and, in general, they have problems in starting a movement. Many of them cannot speak correctly, since they cannot control the vocal chords and the vocal tract.

Levodopa is prescribed to patients, but the amount of treatment should be increased as the illness progresses and it should be provided at the right time during the day, to prevent the freezing phenomenon. It would be beneficial to measure total UPDRS ((Unified Parkinson's Disease Rating Scale) many times during the day in order to adapt the treatment to the specific patient. This means that an automatic way to measure total UPDRS must be developed using simple techniques easily managed by the patient or his/her caregiver.

One possibility is to use patient voice recordings (that can be easily obtained several times during the day through a smartphone) to generate vocal features that can be then used to regress total UPDRS.

Gaussian Process Regression (GPR) was used on the public dataset at [1] to estimate total UPDRS, and the results were compared to those obtained with linear regression, showing the superiority of GPR.

2 Data analysis

The 22 features available in the dataset at [1] are listed in table 1: of these, subject ID and test time were removed, total UPDRS is the regressand. All the remaining 19 features were used as regressors in linear regression, but only 3, namely motor UPDRS, age and PPE, were used in GPR.

The number of points in the dataset is 5875; data are shuffled and the first 50% of the points are used to train the model, 25% of the points are used for the validation and the

1	subject	2	age	3	sex
4	test time	5	motor UPDRS	6	total UPDRS
7	Jitter(%)	8	Jitter(Abs)	9	Jitter:RAP
10	Jitter:PPQ5	11	Jitter:DDP	12	Shimmer
13	Shimmer(dB)	14	Shimmer:APQ3	15	Shimmer:APQ5
16	Shimmer:APQ11	17	Shimmer:DDA	18	NHR
19	HNR	20	RPDE	21	DFA
22	PPE				

Table 1: List of features

remaining 25% are used to test the model performance. Data are normalized using mean and standard deviation measured on the training dataset.

3 Gaussian Process Regression

In GPR, it is assumed that $N-1$ measured datapoints $(\mathbf{x}_k, y_k)$ are available in the training dataset, and that a new input $\mathbf{x}_N$ is present, whose corresponding output y_N has to be estimated.

In the following, $\mathbf{Y}_L = [Y_1, \ldots, Y_L]$ is the L-dimensional random vector that includes the random variables Y_ℓ and $\mathbf{y}_L = [y_1, \ldots, y_L]$ is the L-dimensional vector that stores the measured values of Y_ℓ. Vector $\mathbf{x}_\ell$ stores instead the measured regressors for Y_ℓ. The random variable to be estimated is Y_N, knowing the corresponding regressors $\mathbf{x}_N$, and the training dataset made of $N-1$ measured couples $(\mathbf{x}_\ell, y_\ell)$, $\ell = 1, \ldots, N-1$.

- The $N \times N$ covariance matrix $\mathbf{R}_{Y,N}$ of $\mathbf{Y}_N$ has n, k value:

$$\mathbf{R}_{Y,N}(n,k) = \theta \exp\left(-\frac{\|\mathbf{x}_n - \mathbf{x}_k\|^2}{2r^2}\right) + \sigma_\nu^2 \delta_{n,k}, \quad n,k \in [1,N]$$

- $\mathbf{R}_{Y,N}$ can be rewritten as

$$\mathbf{R}_{Y,N} = \begin{bmatrix} \mathbf{R}_{Y,N-1} & \mathbf{k} \\ \mathbf{k}^T & d \end{bmatrix}$$

  where $\mathbf{R}_{Y,N-1}$ is the covariance matrix of $\mathbf{y}_{N-1}$.

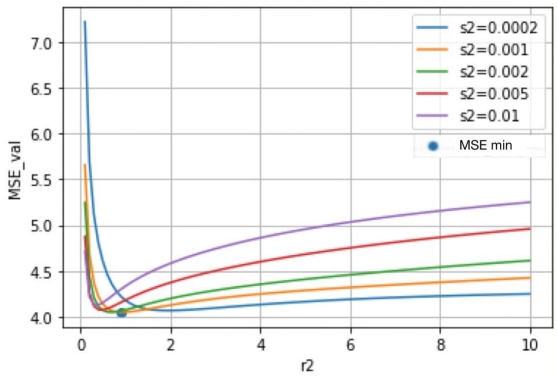
- Then the pdf of Y_N given the measured values $\mathbf{y}$ of $\mathbf{y}_{N-1}$ is

$$f_{Y_N|\mathbf{y}_{N-1}=\mathbf{y}}(z) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z-\mu)^2}{2\sigma^2}}$$

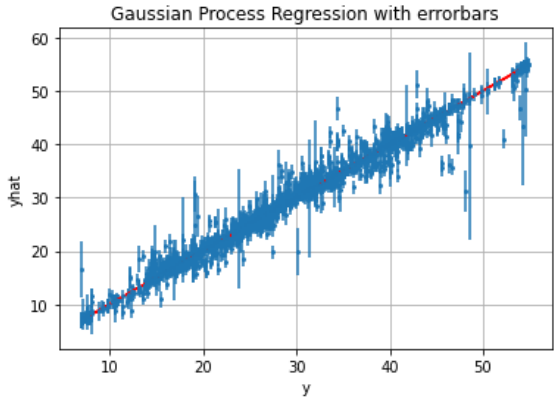
$$\mu = \mathbf{k}^T \mathbf{R}_{Y,N-1}^{-1} \mathbf{y} \tag{1}$$

$$\sigma^2 = d - \mathbf{k}^T \mathbf{R}_{Y,N-1}^{-1} \mathbf{k} \tag{2}$$

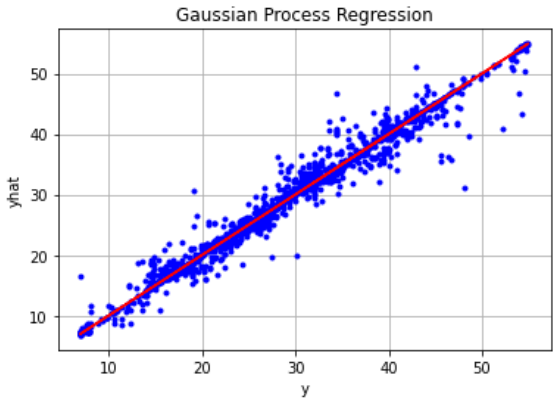
  The point estimation of Y_N is $\hat{y}_N = \mu$.



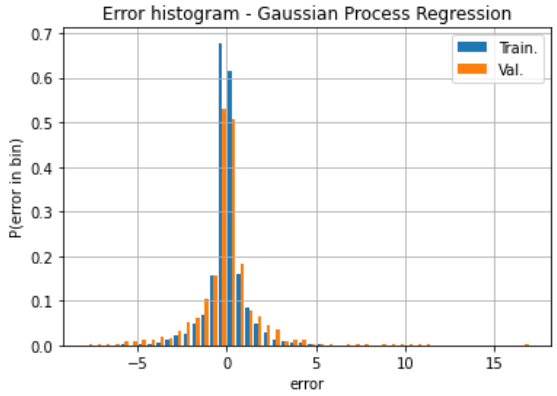
(a) Optimization of $\texttt{r2}=r^2$ and $\texttt{s2}=\sigma_\nu^2$.



(b) $\hat{y}$ versus y with errorbars for the test dataset.



(c) $\hat{y}$ versus y for the test dataset.



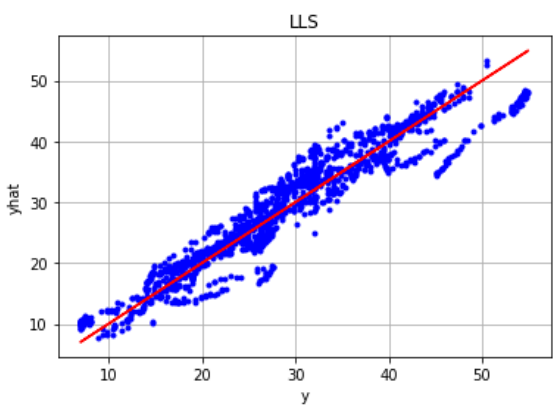
(d) Histogram of $y-\hat{y}$ for training, validation and test datasets.

Figure 1: Gaussian Process Regression results.

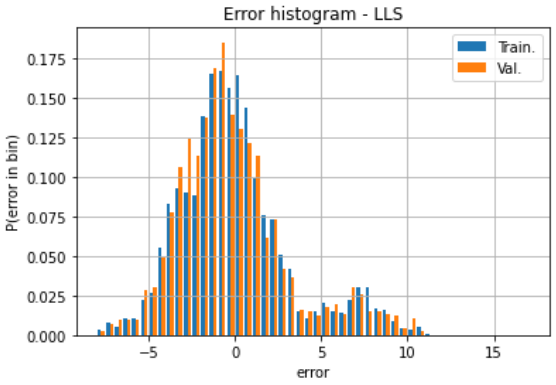
- In the above equations, couples $(\mathbf{x}_\ell, y_\ell)$ for $\ell = 1, \ldots, N-1$ belong to the training dataset, couple $(\mathbf{x}_N, y_N)$ belongs to the test or to the validation dataset.

The model hyperparameters are three: θ, r^2 and σ_ν^2. Since the training dataset stores normalized data, and σ_ν^2 is small, parameter $\theta = \mathbf{R}_{Y,N}(n,n)$ (variance of y_n) was set equal to 1. Hyperparameters r^2 and σ_ν^2 were set to minimize the mean square error $\mathbb{E}\{[y_N - \hat{y}_N]^2\}$ for the validation dataset. In particular, for each point $(\mathbf{x}_N, y_N)$ in the validation dataset, the $N = 10$ closer points in the training dataset were found, a set of possible values for r^2 and σ_ν^2 was tried and the optimum values were found among the considered cases (see Fig. 1a): these optimum values are $r^2_{opt} = 0.9$ and $\sigma^2_{opt} = 0.001$.

Fig. 1c shows $\hat{y}$ versus y whereas Fig. 1b also shows the error bars ($\pm 3\sigma_y$ where σ_y is the denormalized version of σ in (2)). The estimation error histogram is shown in Fig. 1d. Figs. 1b-1d were obtained using r^2_{opt} and σ^2_{opt}.



(a) $\hat{y}$ versus y for test dataset.



(b) Histogram of $y-\hat{y}$ for training, validation and test datasets.

Figure 2: Linear Least Squares results.

4 Linear regression based on Linear Least Squares

The model assumed in linear regression is

$$Y = w_1 X_1 + \ldots + w_F X_F = \mathbf{X}^T \mathbf{w} \tag{3}$$

where Y is the regressand (total UPDRS), $\mathbf{X}^T = [X_1, \ldots, X_F]$ stores the F regressors[1] and $\mathbf{w}^T = [w_1, \ldots, w_F]$ is the weight vector to be optimized. In (3), $Y, X_1, \ldots, X_F$ are all random variables.

Linear Least Squares (LLS) minimizes the mean square error (MSE) and the optimum weight vector $\mathbf{w}$ can be obtained in closed form as:

$$\hat{\mathbf{w}} = \arg\min \mathbb{E}\{(Y - \mathbf{X}^T \mathbf{w})^2\} = \left(\underline{\mathbf{X}}^T \underline{\mathbf{X}}\right)^{-1} \underline{\mathbf{X}}^T \mathbf{y} \tag{4}$$

where $\underline{\mathbf{X}}$ is the matrix that stores the (normalized) training regressor points and $\mathbf{y}$ is the (normalized) training regressand vector. Given $\hat{\mathbf{w}}$, the normalized regressand is estimated as

$$\hat{y}_N = \mathbf{x}_N^T \hat{\mathbf{w}} \tag{5}$$

Figure 2 shows the results obtained with LLS. Note that, to get a meaningful comparison with GPR, the training dataset and test datasets with the two regression models are the same; the validation dataset was only used for GPR, not for LLS regression.

5 Comparison

It is evident, by comparing Figs. 1c and 2a that, with the Parkinson's dataset, Gaussian Process Regression (GPR) is more precise than linear regression, and this is also confirmed by the estimation error histograms in Figs. 1d and 2b.

[1] $\mathbf{X}$ is a column vector and $\mathbf{X}^T$ is its transpose

Table 2 lists the main statistical properties of the estimation error $e = y - \hat{y}$ for the training, validation and test datasets. The mean square error on Test set of GPR is about 1/3 than that of LLS.

	Dataset	Err. Mean	Err. St. dev.	MSE	R^2
LLS	Training	0.0	3.224	10.393	0.984
	Test	-0.170	3.244	10.553	0.984
GPR	Training	0.002	1.04	1.081	0.998
	Test	-0.065	1.981	3.579	0.995

Table 2: Numerical comparison between GPR and LLS.

6 Conclusions

Being the Unified Parkinson's Disease Rating Scale widely used and extensively tested for its clinimetric properties, Regression methods can represent a great relief for medical doctors. Parkinson's is a neurological movement disorder which progresses slowly, therefore it is important to have a way to constantly and objectively (without neurologist's subjectivity) evaluate UPDRS values in a reliable way.

Researches say about 89% of people affected by Parkinson's will have speech and voice symptoms. This observation obtains a fundamental value in comparison to other symptoms such as resting tremor, rigidity, akinesia, bradykinesia (which are due to loss of function of the basal ganglia involved in the coordination of body movement) which are clearly more difficult to be evaluated without a medical doctor intervention. In fact the Perceived Vocal Effort (PPE) is used in this analysis, since it is a parameter specifically thought for people affected by PD (thus very informative) and since parameters related to voice are automatically evaluated by specific software that uses a voice signal as input. Performance by PD subjects on the constant-effort task resembled that by normal adults who were pre-fatigued. Researches results[2] support greater than normal sense-of-effort related to fatigue in PD, and provide preliminary validation of a performance-based physiologic task to assess abnormal sense of effort in this population. Theoretically speaking, features that can not be measured only through voice recording should not be considered to make the analysis more flexible for common people use: in this case, for instance, motor UPDRS should be removed but doing that regression shows much worse results (for instance MSE on test set will be equal to 95.5 for LLS). More generally, ICT could intervene on Speech symptoms, respiration, phonation and articulation.

Furthermore, more measurements during the day are useful to neurologist who can optimize when and how much levodopa the patient should take in order to increase dopamine in the brain and decrease motor dysfunction. A correct and recurrent evaluation is fundamental because as the disease progress, more dopaminergic neurons in the substantia nigra are lost and conversion of levodopa to dopamine decreases; moreover as the movements be-

come slower and slower, levodopa stays more and more in the stomach without reaching the intestine where it should be absorbed.

From the analysis of the comparison between $\hat{y}$ and y (fig. 2a) it is clear that the LLS regression method performs well for y values smaller than 40. For values greater than 40 it can be observed that predicted values are more distant from the real values. This observation also explains the presence of two gaussian-like pdf in the histograms of $e = y - \hat{y}$ (fig. 2b): the one with the higher probabilities is referred to value y<40, in fact errors are close to zero; the other one for errors related to y >40 (except for some outliers observable in the range 20< y <30).

This range is the area in which the better results of Gaussian Process Regression are more evident. Qualitatively, fig. 1c shows clearly that points far from the correct values line are much less than those of LLS.

Quantitatively and more in general, this trend is confirmed by the R^2 values (Table 2) and by the error histograms (figs. 1d and 2b): the GPR histogram shows errors more densely distributed around the mean than the LLS errors. This observation, together with variance values (Table 2), suggests that GPR results are definitely more precise and more reliable. In fact, fig. 1b shows that the ranges of values assumed by $\hat{y}$ with a probability of 99.7% (having used three standard deviations from the mean) almost always include the right y. Furthermore, the error standard deviation on the test set (1.981) is enough smaller than the total UPDRS standard deviation (8.90). However, error means are not zero (but slightly different) since the total UPDRS means in the test subset might not be exactly zero, after normalization. In addition, the GPR method is implemented using less features (3 instead of 19); LLS gives even worse results using only 3 features.

Moreover, the strong similarities of error histograms (test set and training set) for both method show the absence of overfitting.

Fig. 1a shows some examples of how non normalized MSE(r^2) of validation set changes as s^2 changes. In the real analysis, much more s^2 were tried to choose the better parameters.

The goodness of the analyzed methods can also be seen through the lens of Mean Squared Error (Table 2). GPR is clearly the better method as MSE reaches the value of 3.579 for the test set (compared to the LLS MSE of 10.553 for test set). The analyzed values can be translated to an estimation of the regressor errors that are almost always between 3-6 for LLS and between 2-4 for GPR. These values, compared to the total UPDRS scale (regressand), mean that both the results might be accepted by medical doctors, since total UPDRS has mean equal to 36.2 in the test set. However, the GPR regression should be the used method, giving better results.

References

[1] `https://archive.ics.uci.edu/ml/datasets/Parkinsons+Telemonitoring`

[2] `https://www.sciencedirect.com/science/article/pii/S135380200500132X?via%3Dihub`