Ambiguities in Fit-Evaluation for Selector Models

B Twala
Institute for Intelligent Systems
Faculty of Engineering and Built Environment
University of Johannesburg
Johannesburg, South Africa
btwala@uj.ac.za

M.V. Seotlo
Remote Sensing Research Unit
CSIR
Pretoria, South Africa
Johannesburg, South Africa
vseotlo@csir.co.za

***Abstract*—The use of the direct evaluation of the Gaussian Process, using the square exponential function kernel prediction at the given data points is often misleading towards evaluation of the fit, given by the coefficient of determination. The predicted value at the data points when using the Gaussian Process, is almost at all cases equal to the original value. As such, interpretation problems arise when coefficient of determination suggest the model to be a good fit, but visual representations suggest otherwise. We illustrate the difficulties in presenting the coefficient of determination for the Gaussian Process and recommend the use of alternative methods for the evaluation of the predicted value, thus realizing the true function of the coefficient of determination.**

Keywords *- Gaussian process, selector models, goodness-of-fit; coeffient of determination*

I. INTRODUCTION

How does one evaluate the quality for a computational or mathematical model of a Gaussian process? Does the structure implied by the model account for the observed variances and covariances among a set of variables? The answer to the two questions is always important yet so complicated. It is important because mathematics makes it possible to formalize the reaction of the process. Model evaluation is complicated because it involves subjectivity, which can be difficult to quantify [7; 9]. A well-fitting model could still have substantial problems and ambiguities that would typically be ignored by users. In other words, there are no formal methods for how to evaluate the quantitative goodness-of-fit of models to data. To address this complication measures such structural equation modelling [8] and coefficient of determination have been be utilized to deal with this problem. The paper follows the latter strategy.

The coefficient of determination (denoted by R^2 since it is the square of the correlation coefficient) is very useful because it gives the proportion of the variance (fluctuation) in the independent variable that is predictable from the independent variable or put simpler, in determining the degree of linear correlation of variables ("goodness of fit") in regression analysis[5; 10]

The coefficient of determination, R^2, is a good and popular measure of fit which relies on properly defined predictions and data points. Without proper definitions, undesirable misuse and misinterpretation may become frequent. Such issues arise, because, it is assumed that the data points are valid, and their predicted counterparts are the direct evaluation of the chosen model. Furthermore, the use of the model directly towards fit, is generally assumed.

The aim of this discourse is to point out the ambiguity of always evaluating the model directly for the goal of fit evaluation. This is of great importance for functions that are generally known as selector models, examples including the Gaussian Process (with the Square Exponential kernel) amongst many others. These models evaluate to specific values at the given data points, and as such put forward an opportunity for misinterpretation. We present a critique about this matter and recommend the use of alternative means of evaluating the prediction value that is used towards evaluation of fit, by employing the limits near the intended data points. Gaussian Processes have been applied to many engineering and science problems including regression [4], classification [4] and supervised or machine learning problems [3]. In this paper, we follow the regression approach. To this end, the contribution of the paper is restricted to the Gaussian Process Regression using the Square Exponential (with noise) kernel, as they are the most commonly used combination.

The rest of this paper is organized as follows. Section 2 summarizes the coefficient of determination followed by the Gaussian Process and an example illustration in Section 3. Section 4 gives a brief discussion regarding the evaluation of the Gaussian Process via alternative approaches. Finally, the conclusions obtained in this research work are shown in Section 5.

II. COEFFICIENT OF DETERMINATION

Assume a dataset $(X, Y) = \{(X_0, Y_0), \ldots, (X_{\{n-1\}}, Y_{\{n-1\}})\}$, where X is the independent variable, Y is the dependent variable and N is the length of the dataset. A model M, is fitted on the dataset and predicted

values given by $y_* = \{y_*[0], \ldots, y_*[N-1]\}$, where N is the length of y_*.

The measure of fit, the coefficient of determination (R^2), explains the fit as a percentage of the variation in the data that is explain by the chosen model. It defined as.

$$R^2 = 1 - \frac{SSR}{SST} \quad (1)$$

where SSR/SST is the percentage of the sums not accounted for [1], SSR is the sum of the residuals, written as

$$SSR = \sum[(y_*[i] - Y[i])]^2 \quad (2)$$

and SST is the measure of spread (or total sum of squares), given by

$$SST = \sum\left(Y_i - mean(Y)\right)^2 \quad (3)$$

In a standard linear simple regression model, $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$, were β_0 and β_1 are coefficients; Y and X are the regressand and the regressor, respectively, and ε is the error term.

The coefficient of determination is the square of the correlation between predicted scores and actual scores; thus, it ranges from 0 to 1. The value of R^2 closer to 1 indicate a good fit.

III. GAUSSIAN PROCESS

Here we give an overview of the Gaussian Process, and would like refer those interested in detailed explanation to [2] and [3].

The Gaussian Process is collection of random variables, any finite number which have a joint Gaussian distribution [3].

According to [4], a Gaussian Process extends multivariate Gaussian distributions to infinite dimensionality and "generates data located throughout some domain such that any finite subset of the range follows a multivariate Gaussian distribution" [4; 11].

Since a Gaussian Process is completely specified by its mean function and covariance function, and most often with the mean function assumed to be zero [4] mostly for notational simplicity [3]. Thus, relationship between two observations, x and x', is governed by the covariance function, $k(x, x')$ [4].

The mean function and the covariance function are defined as

$$m(x) = E[f(x)], \quad (4)$$

and

$$k(x, x') = E[(f(x) - m(x))(f(x') - m(x'))]. \quad (5)$$

The Gaussian Process is then formulated as

$$f(x) \sim GP(m(x), k(x, x')) \quad (6)$$

with f(x) representing the function value at location x [3], and to account for the noise, we write,

$$g(x) = f(x) + N(0, \sigma_n^2) \quad (7)$$

In the Gaussian Process Regression, the objective is to calculate the value of y at any given x_* given the observations y and this is achieved in the following manner: First, define

$$K = [[k(x_1, x_1) \ldots k(x_1, x_n)], \ldots, [k(x_n, x_1), \ldots k(x_n, x_n)]] \quad (8)$$

$$R = [k(x_*, x_1) \ldots k(x_*, x_n)] \quad (9)$$

[3] and $P = k(x_*, x_*)$, where $k(a, b)$ as the covariance function. Thus y is given by

$$y = RK^{-1}y \quad (10)$$

and the variance given by

$$var(y) = P - RK^{-1}R^T \quad (11)$$

[4, 3].

Each observation y can be thought of as related to an underlying function $f(x)$ through a Guassian noise model

$$y = f(x) + \aleph(0, \sigma_n^2) \quad (12)$$

Now, the Square Exponential (with noise) is utilized in this note.

$$k(x, x') = \sigma_f^2 \exp\left(\left[-\frac{(x - x')^2}{2l^2}\right]\right) + \sigma_n^2 \delta(x, x') \quad (13)$$

where the maximum allowable covariance is defined as σ_f^2. $\delta(x, x')$ is the Kronecker delta function. In the equations above, σ_f , σ_n and l are the hyper-parameters.

IV. STUDY METHODOLOGY

The Square Exponential is a stationary covariance function [3], and as such, the Gaussian Process may be loosely viewed as a selector function. The selector functions are characterized by the property that they evaluate to $g(x)$ in the defined interval $[a, b]$, except at $X[i] \in [a, b]$, where they evaluate to a specific value, $Y[i]$.

The Gaussian Process evaluate to almost $Y[i]$ at $X[i]$, and to y in all other cases. Because the Gaussian Process fits the loose definition of the selector function, it means that $\exists\, \sigma^-$ and σ^+ such that

$$\lim_{\sigma^- \to 0} g_{\sigma^-}(X[i]) \to Y[i] \text{ and} \quad (13)$$

$$\lim_{\sigma^+ \to 0} g_{\sigma^+}(X[i]) \to Y[i], \quad (14)$$

where g_σ is the Gaussian Process parametrized by σ.

Evaluating fit at $X[i]$ within the interval $(X[i] - \sigma^-, X[i] + \sigma^+)$ will give unreliable and at times, an illusion of good fit since the predicted value, $y_*[i]$ will be very close to Y [i]. It is thus better to use information outside of $(X[i] - \sigma^-, X[i] + \sigma^+)$ when evaluating $y_*[i]$.

To achieve the above, we note that $g(X[i] + \sigma^-)$ and $g(X[i] + \sigma^+)$ do not turn to $Y[i]$, thus the proper technique for evaluating y∗[i] simply entails locating $\sigma^-$and σ^+, and using them is some way to get the desired $y_*[i]$, which may then be used to evaluate fit. Algorithm 1 may be utilized towards the goal.

Once the points $A = (X[i] + \sigma^-)$ and $B = (X[i] + \sigma^+)$ are located, the next step entails assuming a smooth transition from A to B. In the simplest case, a linear function going through these points may be defined and evaluated at $X[i]$, thus giving the desired $y_*[i]$ (see figure 2) using algorithm 2.

Using the two algorithms, we are able to evaluate $R^2 = 0.991$ and y∗ as summarized in table 1, compared to $R^2 = 1.0$ when using direct evaluations. Alternatively, $0 < \sigma \ll 1$ may be added to σ^- and σ^+, thus giving extra evaluations outside of $(X[i] + \sigma^-, X[i] + \sigma^+)$. With this extra evaluations, interpolations functions such as the cubic spline

or the Hermite spline may be put to use when there is a belief that the rate of change from the left of A is different to the rate of change from the right of B.
The same argument can be applied to the variance function of the Gaussian Process to achieve a "meaningful" value when evaluating at $X[i]$. This is illustrated in figure 3.

V. CONCLUSION

Evaluating the goodness-of-fit using the coefficient of determination require accurate representation of the model at the recorded observations so that it does not become ambiguous and misleading at times. The direct interpretation of selector-like functions and models such as the Gaussian Process when using stationary kernels (covariance functions) offer opportunity for abuse and erroneous representations.

Our analysis suggests that the direct evaluation of the Gaussian Process and similar functions at the observations may not be used towards fit and such use should be questionable and reconsidered. In fact, it is important for users to bear in mind what precisely is being processed when assessing model fit.

X	***Y***	***Direct y_****	***Proposed y_****
-1.5	-1.6	-1.6	-1.551
-1.0	-1.1	-1.1	-0.939
-0.75	-0.35	-0.35	-0.453
-0.4	0.25	0.25	-0.253
0.25	0.55	0.55	0.515
0	0.9	0.9	0.850

Table 1: Example dataset and evaluation of $\boldsymbol{y_*}$ directly and via the proposed method.

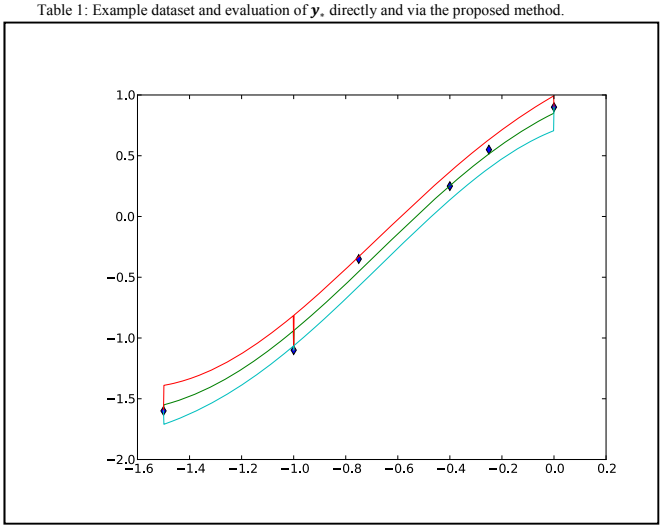

Figure 1. Example model dataset T = { (-1.5,-1.6), (-1.0,-1.1), (-0.75,-0.35), (-0.4,0.25), (-0.25,0.55), (0.0,0.9) }. The parameters for the Gaussian Process {σf , σn, l} = {1.3, 0.3, 1.0}

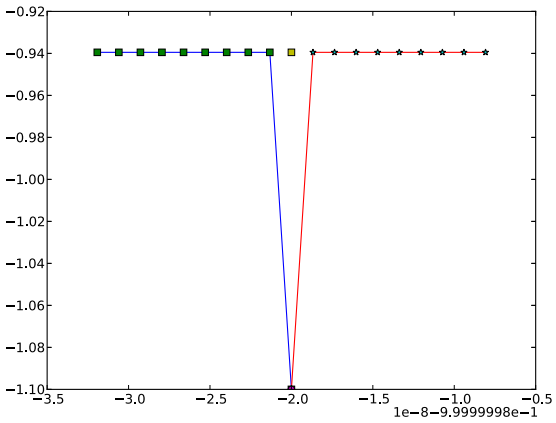

Figure 2. Mean function plot near (−1.0,−1.1). The predicted value (in a yellow squares) is in line with the near linear approximations.

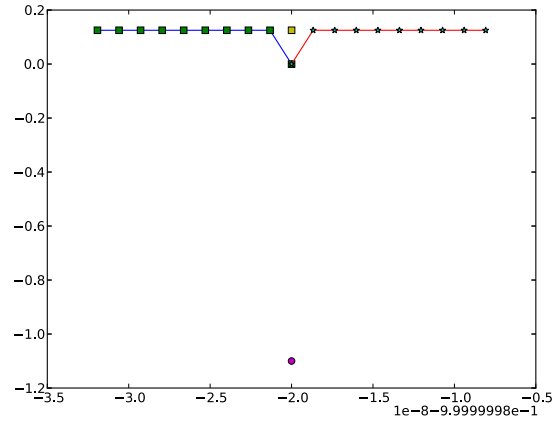

Figure 3. Variance function plot near (−1.0,−1.1). The predicted value (in a yellow squares) is in line with the near linear approximations.

VI. REFERENCES

[1] J. Rawlings, S. Pantula, and D. Dickey, "Applied Regression Analysis: A Research Tool (Springer Texts in Statistics)". Wadsworth & Brooks,1998.

[2] D. J. MacKay, "Introduction to gaussian processes," NATO ASI Series F Computer and Systems Sciences, Vol. 168, pp. 133–166, 1998.

[3] C. E. Rasmussen and C. K. Williams, "Gaussian processes for machine learning," MIT Press, 2006.

[4] M. Ebden, "Gaussian processes for regression:A quick introduction," http://www.robots.ox.ac.uk/mebden/reports/GPtutorial.pdf, 2008 (Accessed in May 2016)

[5] K.V. Mardia, J.T. Kent, and J.M. Bibby, "Multivariate Analysis". Academic Press, 1980.

[6] N.R. Draper, and H. Smith, "Applied Regression Anaysis". John Wiley, 1988.

[7] C.D. Schunn, and D. Wallach, D, "Evaluating goodness-of-fit in comparison of models to data". In W. Tack (Ed.), Psychologie der Kognition: Reden and Vorträge anlässlich der Emeritierung von Werner Tack (pp. 115-154). Saarbrueken, Germany: University of Saarland Press, 2005.

[8] R. Kline, "Principles and Practice of Structural Equation Modeling", (3rd Ed.). Guilford, 2011.

[9] J.J. Hox, "Multilevel Analysis: Techniques and Applications", (2nd Ed) Mahwah, NJ: Lawrence Erlbaum, 2010.

[10] W.J. van der Linden & J. Glas, "Statistical tests of conditional independence between responses and/or response times on test items", Psychometrica, Vol 75 (1), 120-139, 2010.

[11] D. Sivia & Skilling, J. "Data Analysis: A Bayesian Tutorial ", (2nd Ed), Oxford Science Publications.

AUTHORS' BACKGROUND

Your Name	Title*	Research Field	Personal website
M.V Seotlo	Master student	Remote Sensing	
B. Twala	Prof.	Image and Signal Processing	

*This form helps us to understand your paper better, the form itself will not be published.

*Title can be chosen from: master student, Phd candidate, assistant professor, lecture, senior lecture, associate professor, full professor