International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 07 July 2024, pp: 471-474 www.ijaem.net ISSN: 2395-5252

Parametric versus Non-Parametric Models

Dr.Ch.Prasuna
Assistant professor on contract
Department of Statistics
Vs University,Nellore

Date of Submission: 10-07-2024 Date of Acceptance: 20-07-2024

ABSTRACT
In machine learning, a model learns from training data to map a target function. However, the configuration of the function is undetermined. As a result, we test out many machine learning algorithms to effectively determine which models the intrinsic function. To tune the behavior of models for a given task, these models are parameterized. We seek to understand such a process by dissecting parametric and non-parametric models.
Keywords: Parametric approach, Non parametric approach

I. INTRODUCING PARAMETERS

A parameter can be described as a configuration variable that is intrinsic to the model. Model parameters are usually not set manually. Parameters are often mistaken for hyper parameters. Hyper parameters are configuration variables that are external to the model. Unlike parameters, hyper parameters are manually set. The value of a parameter can be approximated from the training data in consideration. After training, the parameters would be used to determine the performance of the model on test data. The model uses them to make predictions. A machine learning model with a set number of parameters is a parametric model. Those without a set number of parameters are referred to as non-parametric. We shall dive deeper into this later. As we will dissect later, the coefficients of a linear regression function are examples of model parameters. Another example is in the form of the coefficients in logistic regression. In a neural network, the weights act as the parameters.

II. PARAMETRIC APPROACH

In the approach ordinal rating tests are generally a small number of points along the ROC curve. In the parametric approach to ROC Curves we assume a theoretical model like normal distribution for the test value in both groups and developed a bi-distribution model. The most popular model is the bi-normal model for which the ROC Curve has a closed form.

This approach makes an ROC curve smoothed and also estimates AUC using Bi-normal, Bi-exponential model and Bi-generalized exponential distribution, when the diagnostic test results are continuous from the D and H normal populations.

Suppose the distributions of test values follow exponential distribution in both D and H groups are di $\sim N(\mu_D, S_D^2)$ and hj $\sim N(\mu H, S_H^2)$ respectively.
Let $a = \{\mu_D - \mu_H\}/\sigma_D$ and $b = \sigma_D/\sigma_H$ and F(.) is the Standard normal distribution. Then the ROC models given by
$\text{ROC}(t) = \Phi(a + b\,\Phi^{-1}(t))$
The AUC is given by
$\widehat{AUC} = \phi\left(\frac{a}{\sqrt{1+b^2}}\right)$
Lasko et. Al (2005) recommended the following guidelines for calculating $\widehat{AUC}$;

i. If the two distributions are not well separated or one of the distributions has greater complexity (then $\widehat{AUC}$ will be less than 0.8) then empirical method can be used.
ii. For the environment given above a better estimate can be obtained with kernel density method.
iii. If the distributions are well separated either empirical method or binormal method can be used.

Suppose the distributions of test values follow exponential distribution in both D and H graphs withλ_H, λ_D respectively. Betinec (2008) has shown that the bi-exponential ROC is given by
$\text{ROC}(t) = t^{\varsigma}$ where $\varsigma = \frac{\lambda_H}{\lambda_D}$
In general the means of the two populations under study are not exactly known and their values are estimated from sample data.
Vishnu vardhan , Sudesh Pundir and Sameera (2012) has shown that bi-exponential model AUC is

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 07 July 2024, pp: 471-474 www.ijaem.net ISSN: 2395-5252

$$AUC = \frac{\lambda_D}{\lambda_H + \lambda_D}$$

The detailed explanation of bi-exponential ROC method for parametric approach given in the Chapter-2.

Consider the Generalized Exponential (GE) distribution given by the density.

$f(x; 1, \lambda, \mu) = (1/\lambda)\, e^{-(x-\mu)/\lambda}$; $(x > \mu; \lambda > 0;)$

Let (λ_H, μ_H) and (λ_D, μ_D) be the parameters of GE distribution in the H & D group respectively. We shown that the ROC curve is of the form $Y(t) = x(t)^{\beta} e^{\left(\frac{\mu_D - \mu_H}{\lambda_D}\right)}$ where $\beta = \frac{\lambda_H}{\lambda_D}$, and μ_H , μ_D , λ_H and λ_D are the location and scale of the test result in the two groups. It can be shown that the AUC of Bi GE model is

$$AUC = \frac{\lambda_D}{\lambda_D + \lambda_H} e^{\left(\frac{\mu_D - \mu_H}{\lambda_D}\right)}$$

III. NON-PARAMETRIC APPROACH

According to Lasko.etal (2005) this approach is to approximate the ROC curve by simply connecting the data points (Sn, 1-Sp), the straight lines, and then calculating the estimated area $\widehat{AUC}$ using the trapezoidal rule. This is referred to as the empirical method. Bamber (1975) has shown that by this method two sample Wilcoxn rank sum statistic and C-Index.

Let $d_1, d_2, \ldots, d_{nD}$ be the test values for n_D diseased subjects and $h_1, h_2, \ldots, h_{nH}$ be the test values for n_H healthy subjects, and define a comparison function $C(d_i, h_j)$ where

$$C(d_i, h_j) = \begin{cases} 1 & \text{if } d_i > h_j \\ 0.5 & \text{if } d_i = h_j \\ 0 & \text{if } d_i < h_j \end{cases}$$

Then the estimated Area is the Average value of the comparison function for all possible pairs of diseased vs. non-diseased subjects is

$$\widehat{AUC} = \frac{1}{n_D n_H} \sum_{i=1}^{n_D} \sum_{j=1}^{n_H} C(d_i, H_j)$$

Where n_D = Number of diseased subjects

n_H = Number of healthy subjects

The advantage of empirical method there is no structural assumptions of the data, and it is used widely applicable. Farraggi anf Reiser found that smoothed curve methods outperformed the competing methods discussed, when the diseased or healthy distribution was a bimodal mixture and the two were poorly separated. The empirical method comes in a close second in performance.

Let x1,x2,…,xm and y1,y2,…,yn are the diseased and healthy populations with Cumulative distribution functions $\widehat{F_m}(t)$ and $\widehat{G_n}(t)$ respectively. The empirical cumulative distribution function is defined for any given value t, to be the observed percentage of sample values less than or equal to t.

Lloyd (1998), Zou et.al(1997) discussed refining the Non- parametric approach to provide a smoothed ROC curve using the Kernel method. Zou etal (1997) proposed the Gaussian Kernel Probability density function $f(t) = F'(t)$ is given by

$$\hat{f}(t) = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{H_x} \phi\left(\frac{t - x_i}{H_x}\right)$$

Where ϕ is the Probability of the standard normal distribution and h_x is bandwidth for x group.

Similarly for subjects y Gaussian Kernel probability function will be

$$\hat{g}(t) = \frac{1}{n} \sum_{i=1}^{m} \frac{1}{H_y} \phi\left(\frac{t - y_j}{H_y}\right)$$

The integration of the Gaussian Kernel estimator of F is

$$\hat{F}(t) = \frac{1}{m} \sum_{i=1}^{m} \Phi\left(\frac{t - x_i}{H_x}\right)$$

Similarly the integration of the Gaussian Kernel estimator of G is

$$\hat{G}(t) = \frac{1}{n} \sum_{i=1}^{m} \Phi\left(\frac{t - y_j}{H_y}\right)$$

Gaussion Kernel Probability density function F (.) or G (.) results in a smoothed estimator of the ROC curve. Lloyd shown that the resulting Kernel estimate of the AUC can be written as

$$AUC = \frac{1}{nm} \sum_{i=1}^{m} \sum_{j=1}^{n} \Phi\left(\frac{x_i - y_j}{\sqrt{H_x^2 + H_y^2}}\right)$$

Zou et.al proposed the band width (H_x) is given by

$H_x = 0.9 \min(S_x, iqr_x/1.34) m^{-1.5}$

Where S_x and iqr_x are the standard deviation and the quartile inter range of the m test results on the diseased sample

Similarly bandwidth (H_y) is given by

$H_y = 0.9 \min(S_y, iqr_y/1.34) m^{-1.5}$

IV. SEMI PARAMETRIC APPROACH

In this approach also there is no need to make the distributional assumptions of the D and H populations like nonparametric approach. This estimates the parameters a and b and the corresponding AUC similar to the parametric approach. So this can be described as a semi parametric approach.

General Linear Model (GLM) is used in the implement of the semi parametric ROC approach.

A binary indicator variable is defined by

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 07 July 2024, pp: 471-474 www.ijaem.net ISSN: 2395-5252

U_{ij} = I $\lfloor X_i \geq Y_J \rfloor$; i= 1,2,….,m; j= 1,2,…..,n for all $m \times n$ possible pairs of diagnostic test results. At each possible pair, the false positive rates t_j is calculated by

t_j = FPR (Y_j); $t_j \in$ T = $\left\{ \frac{1}{n}, \frac{2}{n}, ..., \frac{n}{n} \right\}$

then the ROC curve is constructed parametrically as

$$g\ (\ ROC\ \beta\ (t)\ ) = \sum_{k=1}^{k} \beta_k h_k\ (t)$$

Where g is the specified link function, $h_1, h_2 \ldots h_k$ are bias functions and $\beta_1, \beta_2, \ldots\ldots, \beta_k$, are the unknown parameters. After applying the GLM procedures, the linear model can be derived by using the expectation of the binary variable U_{ij} and the function t_j.

Then the linear mode is given by

$$g\ (\ E(U_{ij})\ ) = \sum_{k=1}^{k} \beta_k h_k\ (t_j)$$

in the above equation, probit link function Φ, $h_1(t_j) = 1$and $h_2(t_j) = \Phi^{-1}(t_j)$ are substituted, the linear model is reduced to the form as

$$E\ (U_{ij}) = \Phi\ (\ \beta_1 + \beta_2\ \Phi^{-1}(tj))$$

Where $\beta_1$and β_2 are the parameter estimates of the linear model i.e. $\hat{\beta}_1$ and $\hat{\beta}_2$.

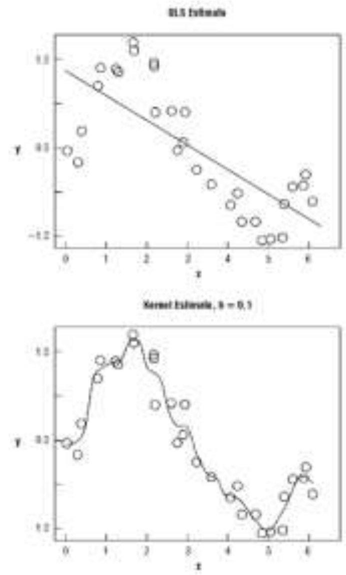
The AUC for the semi parametric model is estimated by using $\hat{\beta}_1$ and $\hat{\beta}_2$ as

$$\hat{AUC} = \Phi \left(\frac{\hat{\beta}_1}{\hat{\beta}_2} \right)$$

And also the variances of $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{AUC}$ are computed by using the bootstrap techniques.

V. PARAMETRIC VS. NON-PARAMETRIC MODELS

Since we can now define both parametric and non-parametric models, we can compare both in this following section. We will then look at the benefits and limitations of both types of models.



Results of parametric and non-parametric regression.

Let's first note that the data points in the two scenarios in the image above are the same.

Recalling the equation that we described earlier, the first image illustrates the mapping function as a linear regression line. This represents a parametric model. We see the consequence of the linear function on the data. A lot of data points are ignored.

The image with a wiggly function represents a non-parametric model. As we mentioned, these algorithms make little to no guesses about the mapping function. As a result, they show greater flexibility and offer a better fit to the data over parametric ones.

VI. BENEFITS

Parametric Models

Simplicity. The methods of parametric algorithms are easier to understand. The interpretability of results is also easier in comparison to non-parametric models.

Training data. Parametric algorithms require less training data than non-parametric ones.

Training speed. They are computationally faster than non-parametric methods. They can be trained faster than non-parametric ones since they usually have fewer parameters to train.

Non-Parametric Models

Performance. Non-parametric models may offer more accurate predictions since they offer a better fit to data than parametric ones.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 07 July 2024, pp: 471-474 www.ijaem.net ISSN: 2395-5252

Flexibility. As shown by the image above, these algorithms provide a good fit for data. They can fit many forms of a function.

Little to no assumptions. Little to no guesses about the mapping function are made. Compared to parametric algorithms, non-parametric algorithms learn more from data. This is because the learning of parametric algorithms may be limited by the assumptions that they make.

VII. LIMITATIONS

Parametric models

Form constraints: Parametric methods constrain an algorithm to a specified functional form.

Fit: These methods do not offer the best fit to data. They are not likely to perfectly match the mapping function.

Complexity: Parametric algorithms offer limited complexity. This means that they are better suited to less complex problems.

Non-Parametric Models

Over fitting: As much as these algorithms tend to fit data better than parametric ones, they are more susceptible to over fitting.

Training data: To give an estimate of the mapping function, these algorithms require much more data than parametric ones.

Speed: Non-parametric algorithms are slower to train since they usually have more parameters to consider for the training.

VIII. CONCLUSTIONS

Through this article, we have introduced parametric and non-parametric models. We have noted a handful of examples of both models. Finally, we also compared how they fit given data points as well as their benefits and limitations.

REFERENCES

[1]. David Farragi and Benjamin Reiser (2002), Estimation of the area under the ROC curve, Statistic in Medicine, 21: 3093 - 3106.

[2]. Lasko, T.A. et al (2005), The use of receiver operating characteristics curves in biomedical informatics, Journal of Biomedical Informatics, 38: 404 – 415.

[3]. Krzanowski W. D and J.Hand (2009), ROC Curves for Continuous data, Monographs on statistics and Applied Probability, CRS Press, Taylor and Francis Group, LLC.

[4]. Zou, K.H and Shapiro, DE (1997), Smooth nonparametric Receiver operating characteristic (ROC) curves for continuous diagnostic tests, Statistics in Medicine, 16: 2143-2156.