

# —QUAL TITULO?? —

Lia Sucupira Furtado, Gabriel Alves das Neves, Gabriel da Rocha Silva e André Luis Dantas Gadelha \*

Universidade Federal do Ceará, Dept. de informática Campus do Pici, Av. Mister Hul Brasil

**Resumo.** Na construção de um modelo preditivo com uma performance ótima é necessário um conjunto de passos como o a análise exploratória, o pré-processamento dos dados, a escolha de um modelo que melhor se adque aos dados e a validação do modelo. Na análise dos dados de solubilidade que será feito nesse relatório todos esses passos serão abordados para chegar no modelo de regressão. O modelo de regressão é

In statistical modeling, regression analysis is a set of statistical processes for estimating the relationships among variables. It includes many techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables (or 'predictors'). More specifically, regression analysis helps one understand how the typical value of the dependent variable (or 'criterion variable') changes when any one of the independent variables is varied, while the other independent variables are held fixed.

In a narrower sense, regression may refer specifically to the estimation of continuous response (dependent) variables,

Regression models involve the following parameters and variables. Familiar methods such as linear regression and ordinary least squares regression are parametric, in that the regression function is defined in terms of a finite number of unknown parameters that are estimated from the data.

## 1 Introdução

## 2 Análise Inicial

Inicialmente, é realizado uma análise inicial dos dados para que eles fiquem prontos para se ter uma melhor otimização ao escolher o modelo preditivo.

### 2.1 Análise Exploratório

Nessa fase, é feito uma análise de cada preditor para entender o seu comportamento singularmente. São calculados os histogramas, a média e o desvio padrão de cada preditor para começar a análise se familiarizando com os dados e o que eles representam.

### 2.2 Pré-processamento

Em seguida, os dados são pré-processados, ou seja, os dados são preparados para construir um modelo em cima desse dados com uma melhor performance, em que os dados não afetem os resultados.

Existem várias formas de pré-processamento, uma delas é a transformação dos dados para reduzir o impacto da obliquidade. Essa transformação foi feita usando o método do Yeo-Johnson, que permite valores de obliquidade zero e negativos.

Depois, foi feita uma análise de correlação entre os pares de preditores. Na figura 1, percebe-se uma relação linear entre vários preditores como com o NumMumBonds em relação ao NumBonds, e ao NumNamHAtoms e ao NumAtoms. Essa relação linear indica uma correlação entre esses preditores. Onde podemos prever que em uma matriz de correlação esses preditores ficariam com uma cor azul bem forte.

---

\*This is an optional funding source acknowledgement.



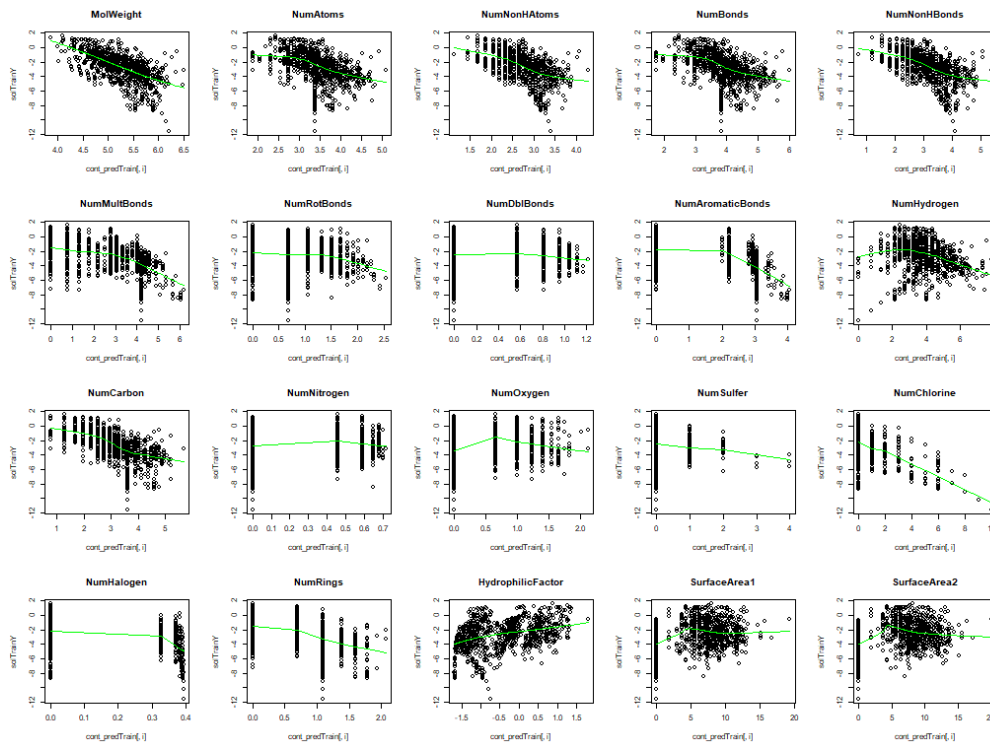


Figura 2: ESANN 2005: Announcement and call for papers.

Os resultados obtidos da 10-fold-cross validation foram que o número de componentes ótimas foram 12 componentes. Pelo gráfico abaixo, podemos ver que o RMSE começa a minimizar nas componentes 7 e 8, e pelo cross-validation se chegou em uma quantidade ótima que minimiza o RMSE.

### 3.1 Tables and figures

Figure 4 shows an example of figure and related caption. Do not use too small symbols and lettering in your figures. Warning: your paper will be printed in black and white in the proceedings. You may insert color figures, but it is your responsibility to check that they print correctly in black and white. The color version will be kept in the ESANN electronic proceedings available on the web.

Table 1 shows an example of table.

| ID | age | weight |
|----|-----|--------|
| 1  | 15  | 65     |
| 2  | 24  | 74     |
| 3  | 18  | 69     |
| 4  | 32  | 78     |

Tabela 1: Age and weight of people.

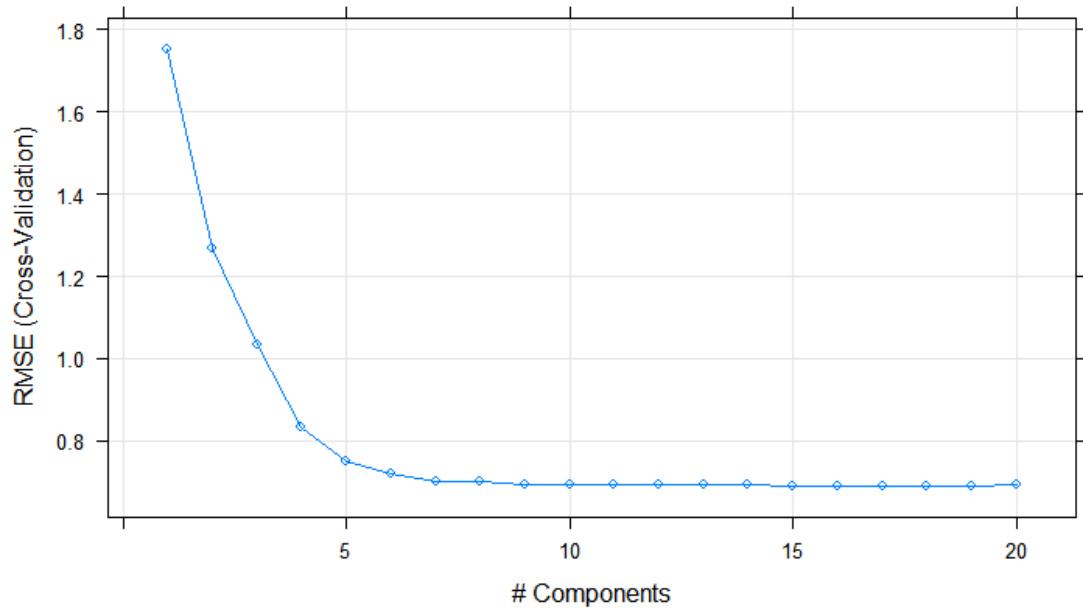


Figura 3: ESANN 2005: Announcement and call for papers.

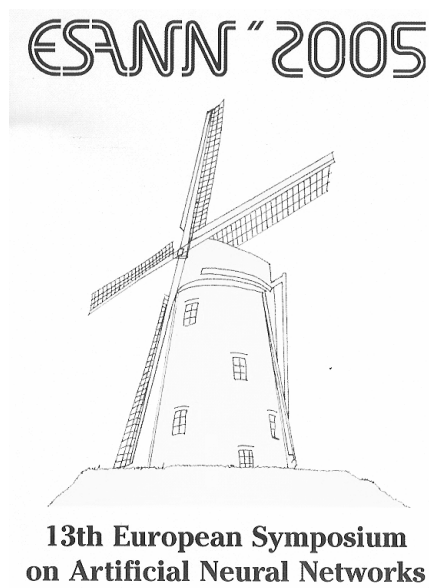


Figura 4: ESANN 2005: Announcement and call for papers.

## 4 Citation

This ESANNV2.tex file defines how to insert references, both for BiBTeX and non-BiBTeX users. Please read the instructions in this file.

## Referências

- [1] S. Haykin, editor. *Unsupervised Adaptive Filtering vol.1 : Blind Source Separation*, John Wiley and Sons, New York, 2000.
- [2] N. Delfosse and P. Loubaton, Adaptive blind separation of sources: A deflation approach, *Signal Processing*, 45:59-83, Elsevier, 1995.
- [3] S. Cruces, A. Cichocki and S. Amari, The minimum entropy and cumulants based contrast functions for blind source extraction. In J. Mira and A. Prieto, editors, proceedings of the 6<sup>th</sup> *international workshop on artificial neural networks* (IWANN 2001), Lecture Notes in Computer Science 2085, pages 786-793, Springer-Verlag, 2001.
- [4] F. Vrins, C. Archambeau and M. Verleysen, Towards a local separation performances estimator using common ICA contrast functions? In M. Verleysen, editor, *proceedings of the 12<sup>th</sup> European Symposium on Artificial Neural Networks* (ESANN 2004), d-side pub., pages 211-216, April 28-30, Bruges (Belgium), 2004.
- [5] J. V. Stone and J. Porrill, Undercomplete independent component analysis for signal separation and dimension reduction. Technical Report, Psychology Department, Sheffield University, Sheffield, S10 2UR, England, October 1997.