**Supervised Learning**

*Alberto González Delgado*

**I have applied different supervised learning algorithms to clinical data from patients diagnosed with colorectal cancer to generate models to predict whether the treatment applied to patients will be effective, based on 21 SNPs features. I studied which parameters allows the best prediction of the target feature.**

**Introduction**

Colorectar cancer is the thid most commonly diagnosed cancer and the second in mortality. Colorectar cancer is a heterogeneous tumour with different genetic alterations. In the last years a better knowledge of the interactions between cancer cells and the immune system has led to immunotherapy and other new approaches (Taunk, T., et al., 2019). However, it is know that the behaviour of certain treatment depends on certain SNPs (Single Nucleotide Polymorphisms). The aim of this study is use Supervised Learning approaches to determine the behavior of the treatment in a patient depending on the SNPs.

Machine learning is a branch of computer science that aims to learn from data to improve performance at various tasks. In healthcare research, machine learning is used to describe and apply computationally approaches to identifying patterns in complex data structures. Supervised learning is used to describe prediction tasks because the goal is to forecast or classify a specific outcome of interest (i.e., the treatment will have good results or not) (Tammy, J., et al., 2020).

**k-Nearest Neighbor (kNN)**

The k Nearest Neighbor (kNN) algorithm is used as a method for big data mining to predict the labels of test data by inducing all training data (Shichao, Z., et al., 2017). It is a non-parametric classification algorithm in which a labelled training dataset is provided where the data points are categorized into various classes, so that the class of an unlabelled data can be predicted. For a new input data, the k nearest neighbors is calculated and the majority label among the neighbouring data decides the label for this new input (Taunk, K., et al., 2019).

**Logistic Regression**

https://dl.acm.org/doi/pdf/10.1145/2990508

**References**

1. Tammy Jiang, Jaimie L. Gradus, Anthony J. Rosellini, 2020; Supervised Machine Learning: A Brief Primer, *Behavior Therapy*, 51(5) pp. 675-687. <https://doi.org/10.1016/j.beth.2020.05.002>.
2. Shichao Zhang, Xuelong Li, Ming Zong, Xiaofeng Zhu, and Debo Cheng, 2017; Learning k for kNN classification. *ACM Trans*. *Intell. Syst.* Technol. 8, 3; 43. <http://dx.doi.org/10.1145/2990508>
3. K. Taunk, S. De, S. Verma and A. Swetapadma, 2019 ;"A Brief Review of Nearest Neighbor Algorithm for Learning and Classification," International Conference on Intelligent Computing and Control Systems (ICCS) , pp. 1255-1260, <https://doi.org/10.1109/ICCS45141.2019.9065747>.
4. Davide Ciardiello, Pietro Paolo Vitiello, Claudia Cardone, Giulia Martini, Teresa Troiani, Erika Martinelli, Fortunato Ciardiello, 2019; Immunotherapy of colorectal cancer: Challenges for therapeutic efficacy*, Cancer Treatment Reviews*, 76, pp. 22-32. https://doi.org/10.1016/j.ctrv.2019.04.003.