

Propuesta de Trabajos Fin de Grado, curso académico 2018-19

PROFESOR: José Luis Torrecilla Noguerales

1.- **TÍTULO:** Selección de variables

Resumen/contenido: Las técnicas de selección de variables permiten reducir la dimensión de los datos manteniendo un nivel alto de interpretabilidad. Existen multitud de aproximaciones al problema dependiendo de los datos, los objetivos o las restricciones de cada problema. El trabajo supondrá una primera aproximación a estas metodologías estudiando sus principales elementos, posibles marcos teóricos y algunos de los métodos más utilizados. Si el estudiante está interesado se podrá complementar el trabajo con simulaciones en problemas de clasificación con datos genéticos en R o MATLAB.

Bibliografía/referencias:

- Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of machine learning research*, 3(Mar), 1157-1182.
- Brown, G., Pocock, A., Zhao, M. J., & Luján, M. (2012). Conditional likelihood maximisation: a unifying framework for information theoretic feature selection. *Journal of machine learning research*, 13(Jan), 27-66.
- Peng, H., Long, F., & Ding, C. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on pattern analysis and machine intelligence*, 27(8), 1226-1238.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). New York: springer.

2.- **TÍTULO:** Random Forest

Resumen/contenido: Las combinaciones de expertos (ensembles) son omnipresentes en los sistemas de aprendizaje automático modernos. Random Forest es el máximo exponente de los ensembles basados en árboles. En este trabajo se estudiarán sus elementos básicos: árboles de clasificación y regresión (CART), remuestreo bootstrap y la metodología bagging. A elección del estudiante se podrán incluir simulaciones con datos reales en R o MATLAB.

Bibliografía/referencias:

- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). New York: springer.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- Breiman, L. (2017). *Classification and regression trees*. Routledge.
- Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2), 123-140.
- Díaz-Uriarte, R., & De Andres, S. A. (2006). Gene selection and classification of microarray data using random forest. *BMC bioinformatics*, 7(1), 3.

*Ambos temas podrían servir para un trabajo en grupo de 2 estudiantes.