

Implementación de algoritmo de submuestreo para *Support Vector Regression*

Jhoan Stiven delgado Villareal

Las Máquinas de Vectores de soporte Vectorial es un algoritmo de aprendizaje supervisado para clasificación y regresión propuesto por *Vladimir Vapnik* y sus colegas en los años 90 [2] . Este algoritmo se basa en la resolución de un problema de optimización cuadrática. Ha mostrado resultados impresionantes cuando se trata de clasificar o realizar predicciones de datos lineales y no lineales gracias a sus funciones *kernel* que le ayudan a generalizar datos más complejos. A pesar de los impresionantes resultados, este algoritmo no es adecuado para grandes conjuntos de datos porque se vuelve inmanejable desde el punto de vista computacional y requiere demasiado tiempo al realizar el entrenamiento debido a que hay que resolver un problema de optimización cuadrática de gran escala, donde el *Hessiano* es grande y denso. Proponemos una implementación extendida y novedosa de un algoritmo de submuestreo (originalmente propuesto por [1]) enfocándonos en regresión de vectores de soporte (SVR) y utilizando el lenguaje de programación Python junto con optimización de hiperparámetros (Bayesian Optimization). Comparamos diferentes métricas como RMSE, MAE, R2 (con respecto a la implementación original [3]) y obtuvimos resultados interesantes, haciendo que nuestra propuesta de implementación sea hasta 67.5 veces más rápida que el algoritmo SVR original.

Resumen Bibliografía:

[1] Camelo, S. A., González-Lima, M. D., & Quiroz, A. J. (2015). Nearest neighbors methods for support vector machines, *Annals of Operations Research*, 235(1), 85-101.

[2] Vapnik, V., Golowich, S. E., & Smola, A. (1997). Support vector method for function approximation, regression estimation, and signal processing. *Advances in neural information processing systems*, 281-287.

[3] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of machine Learning research*, 12, 2825-2830.