World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:16, No:10, 2022

Closest Possible Neighbor of a Different Class: Explaining a Model Using a Neighbor Migrating Generator

Authors : Hassan Eshkiki, Benjamin Mora

Abstract : The Neighbor Migrating Generator is a simple and efficient approach to finding the closest potential neighbor(s) with a different label for a given instance and so without the need to calibrate any kernel settings at all. This allows determining and explaining the most important features that will influence an AI model. It can be used to either migrate a specific sample to the class decision boundary of the original model within a close neighborhood of that sample or identify global features that can help localising neighbor classes. The proposed technique works by minimizing a loss function that is divided into two components which are independently weighted according to three parameters α, β, and ω, α being self-adjusting. Results show that this approach is superior to past techniques when detecting the smallest changes in the feature space and may also point out issues in models like over-fitting.

Keywords : explainable AI, EX AI, feature importance, counterfactual explanations
Conference Title : ICMLA 2022 : International Conference on Machine Learning and Applications
Conference Location : Dubai, United Arab Emirates
Conference Dates : October 13-14, 2022