

# Postprocesamiento geométrico de espacios de embeddings para clasificación de textos con Voronoi

Johan Argenis Franco Rogel\*

**Resumen**—Este artículo estudia si una simple corrección geométrica aplicada después de la extracción de incrustaciones mejora la separabilidad de clases en la clasificación de texto. Se evalúan seis familias de incrustaciones bajo dos condiciones, original y geoméricamente mejorada, utilizando tres paradigmas: clasificación de Voronoi inducida por centroide, máquinas de vectores de soporte y agrupamiento no supervisado. El estudio abarca un corpus multiclase en español, un conjunto de datos de referencia de emociones de grano fino y un conjunto de datos de referencia de noticias. Demostramos que aumentar la isotropía puede destruir sistemáticamente la estructura discriminativa en ciertos regímenes. En el corpus en español, la corrección mejora consistentemente la clasificación de Voronoi y la clasificación supervisada. En el conjunto de datos de referencia de emociones, ayuda principalmente a Voronoi mientras que SVM permanece prácticamente sin cambios. En el conjunto de datos de referencia de noticias, degrada Voronoi, SVM y el agrupamiento en todos los modelos. El artículo también contrasta estos resultados con trabajos previos basados en métodos clásicos o débilmente supervisados y argumenta que la regularización geométrica debe tratarse como una intervención dependiente del conjunto de datos en lugar de un paso de preprocesamiento universalmente válido.

**Palabras clave**— Embeddings textuales, isotropía, anisotropía, Voronoi, clasificación, AG News, GoEmotions.

## Geometric Post-processing of Embedding Spaces for Text Classification with Voronoi

**Abstract**—This paper studies whether a simple geometric correction applied after embedding extraction improves class separability in text classification. Six embedding families are evaluated under two conditions, original and geometrically improved, using three paradigms: centroid-induced Voronoi classification, support vector machines, and unsupervised clustering. The study covers a Spanish multi-class corpus, a fine-grained emotion benchmark, and a news topic benchmark. We show that increasing isotropy can systematically destroy discriminative structure in certain regimes. In the Spanish corpus, the correction improves Voronoi and supervised classification consistently. In the emotion benchmark, it mainly helps Voronoi while leaving SVM nearly unchanged. In the news benchmark, it degrades Voronoi, SVM, and clustering across all models. The paper also contrasts these results with prior work based on classical or weakly supervised methods and argues that

Manuscript received on 11/02/2026, accepted for publication on 09/04/2026. Corresponding author is Johan Argenis Franco Roge (m24ce052@cenidet.tecnm.mx).

The author is with Centro Nacional de Investigación y Desarrollo Tecnológico, Cuernavaca, Morelos, México.