Avances de actividades:

Se utilizo kmeans, KernelKMeans, apclusterK y Robust manifold NMF en corel 2500K. Se midieron medidas de desempeño como pureza y entropía en cada uno de los algoritmos. Se corrigieron los posibles capítulos que tendrá el paper a presentar agregando la información referente a M-Estimators.

Se encontraron 9 bases de datos a usar, 5 de ellas en imágenes y 4 de datos. Estas se encontraron en el paper Robust Manifold NMF, mostrando un posible framework de comparación para nuestros algoritmos.

Se encuentra una diferencia clara en el uso de Robust Learning, entre los autores, uno definiéndolo como algo propio de los algoritmos, otros definiéndolo a partir de Estadística Robusta.

Se encontraron tres nuevas medidas para poder comparar clustering robusto con otros métodos utilizados en la literatura, por un lado se encontró Normalized Mutual Information, Purity and Clustering Accuracy.

Resultados

Se encontró que el uso de KernelKMeans comparado con KMeans en corel 2500 K, hace estimación robusta de la media con esto KernelKMeans no aumenta significativamente su sesgo al aumentar la contaminación. Se encontró que al utilizar un kernel gaussiano en apcluster, el aumento en la contaminación no varía significativamente el parámetro de preferencia para un k de cluster dado.

Al utilizar KernelKMeans, KMeans y apclusterK aumentando la contaminación, tomando primero 7,10 y 20 cluster se ve como la diferencia entre la pureza no es significativa (esto hay que comprobarlo mejor).