

Banco de Dados 2

Análise de resultados kmeans plpgsql

Felipe Augusto Arruda 1948423

João Marcelo Tozato 1913310

Vinicius Ribeiro Furlan 1913409

Curso: Engenharia de Computação

Prof. Pedro Bugatti

Discussão

A partir do agrupamento original do dataset *IRIS*, que pode ser visualizado no arquivo [*irisOriginal.pdf*](#) foi possível compará-lo com diversas abordagens do algoritmo *kmeans*, seja variando o número de iterações desejadas ou o próprio número k , que é o número de agrupamentos desejados. Nota-se que, ao executar o algoritmo com um número de iterações igual a 50 e k igual a 3 ([*irisPLPGSQL_50iteracoes_k=3.pdf*](#)), foi possível ver uma certa similaridade com o agrupamento ideal, porém, ao aumentar este número para 100 iterações ([*irisPLPGSQL_100iteracoes_k=3.pdf*](#)) o resultado ficou mais condizente com o original. Já ao aumentar o número de agrupamentos k fica claro que há uma queda no desempenho do algoritmo, visto que não existem mais do que $k = 3$ agrupamentos, então cada novo agrupamento fora este k ótimo gera um resultado não representativo dos dados reais. Estes resultados podem ser visualizados nos arquivos: [*irisPLPGSQL_100iteracoes_k=4.pdf*](#), [*irisPLPGSQL_100iteracoes_k=5.pdf*](#), [*irisPLPGSQL_100iteracoes_k=6.pdf*](#).