

Comparação de desempenho entre implementações do KNN

João Augusto Dias Neto, Rafael Rodrigues de Souza

1 Introdução

Este relatório tem como objetivo comparar a eficiência e a precisão de duas implementações do KNN: uma versão manual desenvolvida do zero e a versão disponível na biblioteca *sklearn*. Para medida de comparação, os testes em ambas as implementações foram avaliadas utilizando a mesma base de dados, onde 75% dos dados foram utilizados para treinamento e 25% para teste. Os resultados brutos como a matriz de confusão e as métricas brutas de cada execução podem ser encontradas no arquivo do jupyter notebook.

2 Dados Coletados

k	Precisão (%)	Revocação (%)	Acurácia (%)
1	0.75	0.74	0.74
3	0.80	0.79	0.79
5	0.77	0.74	0.74
7	0.75	0.74	0.74
1 (<i>sklearn</i>)	0.75	0.74	0.74
3 (<i>sklearn</i>)	0.80	0.79	0.79
5 (<i>sklearn</i>)	0.75	0.74	0.74
7 (<i>sklearn</i>)	0.75	0.74	0.74

Table 1: Comparação de precisão, revocação e acurácia para diferentes valores de k entre a implementação manual e a do *sklearn*.

k	Tempo (s)	Memória (MiB)
1	944 ms	152.84 MiB
3	897 ms	153.25 MiB
5	960 ms	153.47 MiB
7	1.010 ms	153.48 MiB
1 (<i>sklearn</i>)	906 ms	153.74 MiB
3 (<i>sklearn</i>)	975 ms	153.76 MiB
5 (<i>sklearn</i>)	947 ms	153.79 MiB
7 (<i>sklearn</i>)	1.020 ms	153.82 MiB

Table 2: Comparação de tempo de execução e uso de memória para diferentes valores de k entre a implementação manual e a do *sklearn*.

3 Discussão dos resultados

Em termos de precisão, revocação e acurácia, ambas as implementações demonstraram desempenho semelhante. Entretanto, a implementação do *sklearn* teve um desempenho ligeiramente superior com $k = 3$, alcançando uma precisão de 80% enquanto a implementação manual ficou em 75%. Isso pode ser explicado devido a presença de alguma otimização no algoritmo do *sklearn* como a validação cruzada por exemplo.

Quando analisamos o tempo de execução e o uso de memória, os resultados indicam que a implementação manual utiliza espaço de memória semelhante, porém é geralmente mais lenta do que a versão do *sklearn*. Isso pode ser explicado pela eficiência das bibliotecas do *sklearn*, que são implementadas em C/C++ visando o ganho de performance.