

Recuperação inteligente da informação

Relatório atividade P-1

André Pacheco

Doutorado em Ciência da Computação
Programa de pós-graduação em informática
Universidade Federal do Espírito Santo

Conteúdo

Recuperação inteligente da informação Relatório atividade P-1	1
1 Introdução	1
2 Resultados de classificação	1

1 Introdução

A atividade P-1 tem como objetivo a realização da classificação da famosa base de dados da IRIS ¹ utilizando o algoritmo k-vizinhos mais próximos (KNN) (Fukunaga and Narendra, 1975), o modelo probabilístico e o híbrido entre os dois.

2 Resultados de classificação

A base de dados foi executada 30 vezes para cada um dos modelos. O valor de k do KNN foi escolhido de acordo com a atividade anterior, portanto, 11. O modelo probabilístico se utiliza a distribuição normal para estimar probabilidade de novas entradas de acordo com a classe. Os resultados para cada modelo é apresentado na Tabela 1.

Tabela 1. Acurácia de classificação da IRIS para ambos os modelos

Modelo	Média (%)	Desvio (%)
KNN	96.37	2.91
Probabilístico	95.89	2.96
Híbrido	54.44	7.79

¹ <https://archive.ics.uci.edu/ml/datasets/iris>

De acordo com os resultados apresentados, as acurácias do KNN e do modelo probabilístico estão dentro do mesmo desvio padrão. Portanto, os modelos são similares. Já a metodologia híbrida apresenta um resultado bem abaixo dos demais mostrando-se ineficaz para a classificação desta base de dados. Como a base de IRIS é fácil de ser classificada, para uma conclusão mais embasada, seria necessário testá-los para mais conjuntos de dados de características diferentes.

Bibliografia

- Friedman, J., Hastie, T., Tibshirani, R., 2001. The elements of statistical learning. volume 1. Springer series in statistics New York.
- Fukunaga, K., Narendra, P.M., 1975. A branch and bound algorithm for computing k-nearest neighbors. IEEE Transactions on Computers 100, 750–753.
- Salton, G., Wong, A., Yang, C.S., 1975. A vector space model for automatic indexing. Communications of the ACM 18, 613–620.