

Combinação de Classificadores - Lista 1 - Geração de Classificadores

Nome: Pedro Diamel Marrero Fernández.

QUESTÃO 1,2

Para validar os resultados obtidos pelos métodos *Bagging* e *Random Subspace* foram utilizados os classificadores instáveis *Nieves Bayes* (**Bayes**) e *Árvore de decisão* (**DT**), e o classificador estável **KNN**. Usou-se a base de dados *Breast Cancer Wisconsin* (Original) da *UCI Machine Learning Repository* [1]. Em todos os casos utilizou-se *10-Fold Cross Validation*. Foram empregados 100 classificadores no ensemble. No caso de *Random Subspace* usou-se subespaços de tamanho 5.

Resultados

Bagging: Como observa-se na Fig. 1, em todos os casos, o desempenho do Bagging supera o desempenho dos classificadores individuais. Os melhores resultados são alcançados para os classificadores DT e Bayes, utilizando entre 30 e 50 classificadores. Isso ocorre porque os mesmos são classificadores instáveis, o que é propício para o bom desenvolvimento do método.

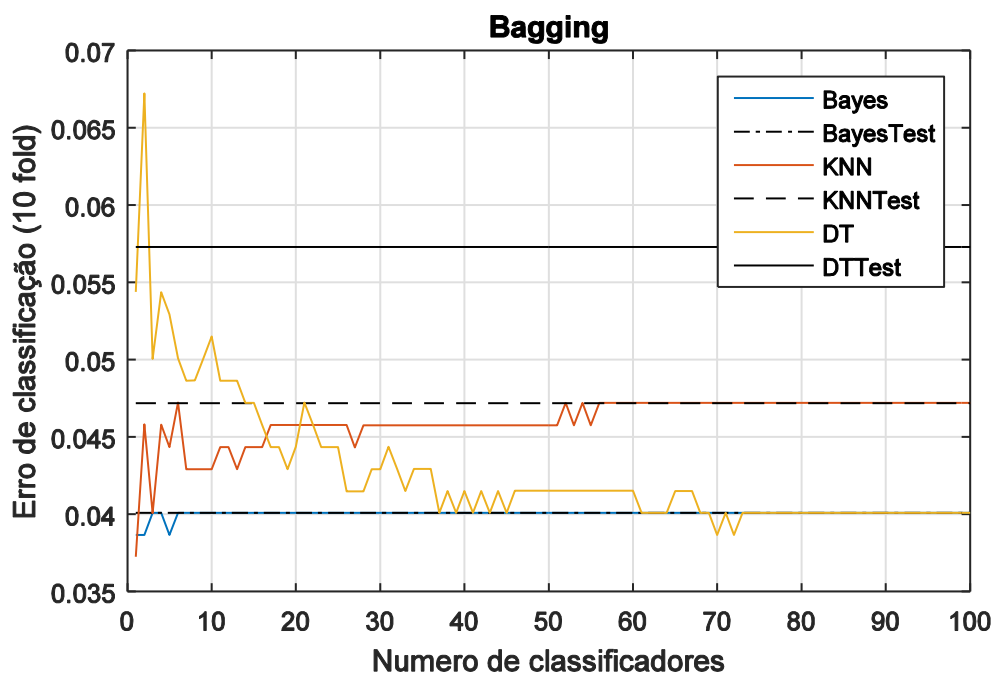


Figura 1. Erros de classificação obtidos usando Bagging.

Random Subspace: A Fig. 2 mostra que os resultados obtidos com Random Subspace superam os resultados dos classificadores individuais. Os menores erros são obtidos para os classificadores DT e KNN.

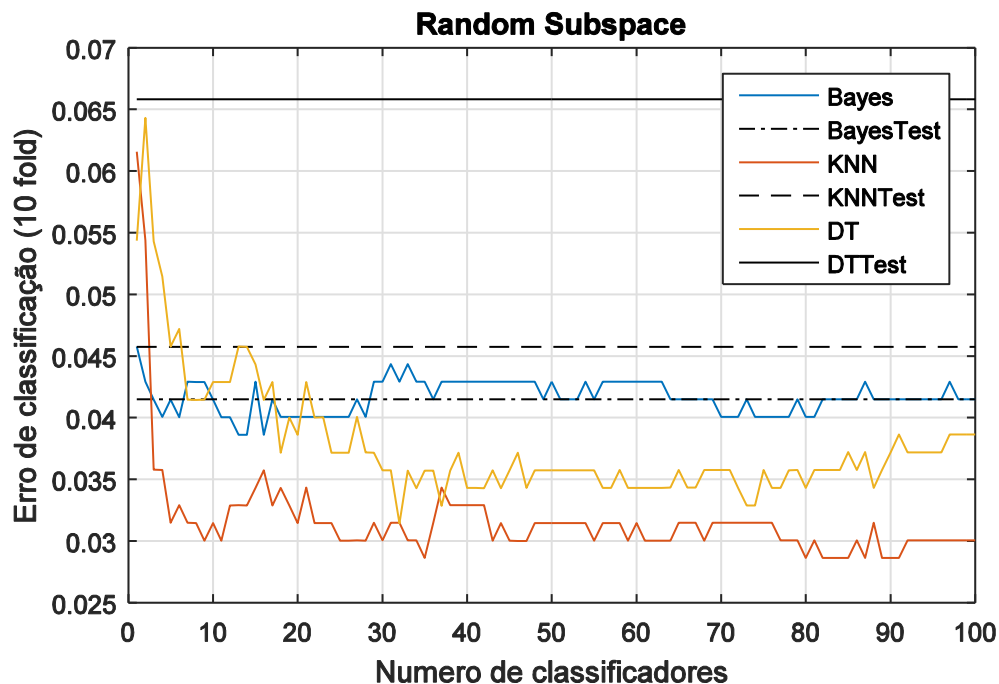


Figura 2. Erros de classificação obtidos usando Random Subspace.

Em geral, observou-se que o Random Subspace tem melhor desempenho do que o Bagging neste banco de dados.

QUESTÃO 3

De acordo com [2] faz sentido aumentar o espaço de características, porque nos espaços de baixa dimensionalidade, as melhorias deste método não são significativas. Para aumentar o espaço de atributos, tenta-se gerar características adicionais através de funções como soma, diferenças e produtos entre as características. Em [3], é mostrado que com o aumento da redundância nos dados não são afetados os resultados do método. Em [5] propõe-se uma estratégia baseada em medidas de dissimilaridade entre objetos para criar novas dimensões.

Referências

- [1] Lichman, M. (2013). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.
- [2] T. K. Ho, "The random subspace method for constructing decision forests," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 20, no. 8, pp. 832–844, 1998.

- [3] Skurichina, M., & Duin, R. P. (2001). Bagging and the random subspace method for redundant feature spaces. In *Multiple Classifier Systems* (pp. 1-10). Springer Berlin Heidelberg.
- [4] L. I. Kuncheva, *Combining pattern classifiers: methods and algorithms*. John Wiley & Sons, 2004.
- [5] E. Pekalska, M. Skurichina, and R. P. W. Duin. Combining Fisher linear discriminant for dissimilarity representations. In J. Kittler and F. Roli, editors, *Multiple Classifier Systems*, Vol. 1857 of *Lecture Notes in Computer Science*, Cagliari, Italy, 2000, Springer, pp. 230–239.