



Aprendizagem de Máquina Relatório Trabalho 2

Acadêmicos: Bianca Carvalho de Oliveira e Gabriela Marim de Souza

Data: 22/04/2019

Descrição do que é o trabalho

Este trabalho tem como objetivo comparar o comportamento, em termos de acurácia, dos classificadores KNN (K Vizinhos mais próximos), Árvore de Decisão, Naive Bayes, Máquina de Vetor de Suporte (SVM) e Multilayer Perceptron (MLP), baseados em diferentes conceitos sobre uma mesma base de dados,

Descrição do conjunto de dados

O conjunto de dados (.arff) utilizado durante os experimentos consiste em uma base chamada Banana, que é composta por 2000 instâncias, as quais são divididas em duas classes, 1 e 2. Onde cada classe é composta de 1000 instâncias e cada instância é composta por 2 atributos, todos do tipo float.

Descrição passo-a-passo do experimento

O experimento foi implementado na linguagem JAVA com o apoio da biblioteca Weka. No primeiro momento carregou-se o implementou-se a divisão estratificada da base Banana, que foi dividida em três subconjuntos mutuamente exclusivos: treino que possui 50% do tamanho da base, validação que possui 25% do tamanho da base e teste que assim como a validação possui 25% do tamanho da base.

A separação das instâncias para cada subconjunto foi feita de maneira totalmente aleatória usando-se um número randômico para a semente de geração de números aleatórios.

Após a separação dos conjuntos foi realizada a implementação dos diferentes tipos de classificadores e escolhido, através do conjunto de validação, os melhores parâmetros para cada classificador obtendo se os seguintes parâmetros como resultado:

- KNN: obteve se o melhor K igual a 13, e a métrica de distância escolhida foi a não ponderada.
- Árvore de decisão: sem poda
- SVM: melhor C igual a 4.0 com Kernel polinomial
- MLP:
 - Taxa de aprendizagem igual a 0,1
 - Número de épocas para treino: 500
 - Número de camadas: 5 (mais diminui a média)

Avaliação dos Experimentos

Depois de implementadas a leitura, divisão de base e escolha do melhor parâmetros para cada classificador, foram realizadas vinte iterações do algoritmo. Onde a cada execução as instâncias que formam cada subconjunto (treino, validação e teste) foram sorteadas novamente. Os resultados obtidos após as vinte execuções são apresentados na Tabela 1 a seguir.

A semente para a geração dos números aleatórios para a divisão da base utilizada foi 10.

Repetição	KNN	Árvore de decisão	Naive Bayes	SVM	MLP
1	0.978	0.97	0.794	0.87	0.978
2	0.984	0.968	0.792	0.85	0.974
3	0.984	0.956	0.746	0.832	0.98
4	0.974	0.95	0.792	0.844	0.968
5	0.984	0.968	0.798	0.858	0.982
6	0.984	0.954	0.778	0.834	0.978
7	0.974	0.962	0.818	0.85	0.974
8	0.988	0.956	0.798	0.866	0.976
9	0.972	0.97	0.806	0.858	0.972
10	0.968	0.966	0.796	0.858	0.967
11	0.97	0.966	0.78	0.848	0.964
12	0.968	0.948	0.768	0.822	0.968
13	0.982	0.968	0.814	0.882	0.982
14	0.982	0.974	0.772	0.844	0.986

15	0.972	0.968	0.772	0.84	0.972
16	0.968	0.972	0.814	0.856	0.978
17	0.986	0.976	0.79	0.86	0.982
18	0.99	0.966	0.8	0.844	0.976
19	0.988	0.99	0.79	0.858	0.98
20	0,98	0.96	0.802	0.858	0.976
Acc (DP)	0.979	0.966	0.791	0.8516	0.976
Desvio Padrão	0.008	0.009713	0.01765	0.01388	0.006

Tabela 1: Resultados

Análise Resultados

Após a realização dos testes de Kruskal-Wallis com 5% de significância, e Mann-Whitney com 5% de significância foi possível confirmar que os classificadores KNN e MLP apresentaram resultados significativamente melhores que os demais classificadores, entretanto entre o KNN e o MLP não há uma diferença significativa.