

Aluno: Deivison Venicio Souza

Lab9: Seleção de Atributos com Algoritmos Genéticos

Considere as bases de treinamento/validação (20000 exemplos) e teste (58646 exemplos) de dígitos manuscritos. Realize a seleção de atributos em modo "Wrapper", e responda as seguintes questões. Escolha dois classificadores de treinamento rápido para realizar seus experimentos. Para cada classificador execute o experimento 10 vezes.

- 1) Qual é o desempenho (base de validação) e a quantidade de características média de características selecionadas?

R: Foram testados 2 classificadores, sob um experimento de 10 vezes. O desempenho médio do classificador LDA na base de validação foi de 0,9576 e, sendo selecionadas em média 80 características. Já para o Perceptron o desempenho médio observado foi de 0,9478, com seleção de 72 características, em média.

- 2) Qual é a diferença média de desempenho entre a base de validação e a base de testes?

R: O desempenho do classificador LDA na base de validação foi de 0,9576, enquanto na base de teste foi de 0,9185. Para o Perceptron o desempenho foi de 0,9478 na validação e de 0,9060 na base de teste, em média.

- 3) Na base de testes, qual é a diferença de desempenho utilizando o subconjunto de características selecionadas e o conjunto original de 132 atributos?

R: Os classificadores (LDA e Perceptron) mostraram desempenho similares usando o subconjunto de características selecionadas e o conjunto original de 132 atributos. Ou seja, o mesmo desempenho foi alcançado usando-se menos features pelos classificadores. Usando-se o conjunto original (132 features) o classificador LDA teve desempenho médio de 0,9271, enquanto após a seleção de características o desempenho foi de 0,9185, em média. De modo similar, o Perceptron teve desempenho de 0,9059 usando todas as features, e após seleção de características seu desempenho foi ficou em 0,9060.

- 4) Nos seus experimentos, existem características que nunca são selecionadas? Se sim, quantas?

R: Não. Considerando as 10 execuções realizadas em algum momento, pelo menos uma vez, uma determinada característica foi selecionada.

Analisando seus resultados, é possível concluir que ambos os classificadores utilizam a mesma quantidade de características e alcançam o mesmo desempenho médio nas bases de validação e teste?

R: Não. O Classificador LDA usou em média 80 features e teve desempenho de 0,9576. Por outro lado, o Perceptron usou 72 características, em média, alcançando desempenho de 0,9478. Portanto, os classificadores não usam a mesma quantidade de features, porém, ainda sim possuem desempenho semelhante. Isso garante relativa vantagem ao Perceptron pelo uso de menos características.

1) Análise Discriminante Linear - LDA

Experimento	Análise Discriminante Linear - LDA						
	Min	Max	std	Rec Rate (validation)	Nb of Features	Performance Using all features (test set)	Performance Using best individual (test set)
1	0.9596	0.9608	0.00026	0.9608	83	0.9281	0.9211
2	0.9597	0.9618	0.00052	0.9604	78	0.9287	0.9224
3	0.9604	0.9614	0.00032	0.9614	78	0.9296	0.9192
4	0.9604	0.9614	0.00025	0.9613	93	0.9275	0.9145
5	0.9586	0.963	0.00112	0.9630	83	0.9255	0.9232
6	0.9502	0.9610	0.00237	0.9502	80	0.9255	0.9216
7	0.9508	0.9567	0.00133	0.9567	77	0.9259	0.9197
8	0.9469	0.9592	0.00296	0.9572	79	0.9267	0.9161
9	0.9492	0.9601	0.00249	0.9492	75	0.9271	0.9167
10	0.9484	0.9576	0.00202	0.9558	76	0.9265	0.9107
Média	0.9544	0.9603	0.0014	0.9576	80	0.9271	0.9185

2) Perceptron

Experimento	Perceptron						
	Min	Max	std	Rec Rate	Nb of Features	Performance Using all features	Performance Using best individual
1	0.8825	0.9369	0.0133	0.9318	70	0.8894	0.8853
2	0.9341	0.951	0.0037	0.9500	67	0.9056	0.9169
3	0.8492	0.9503	0.0220	0.9452	69	0.9070	0.9076
4	0.9316	0.9548	0.0067	0.9516	75	0.9125	0.9060
5	0.8686	0.9411	0.0158	0.9374	73	0.9340	0.8892
6	0.8554	0.9504	0.0208	0.9504	72	0.8965	0.9181
7	0.9569	0.9569	1.5E-08	0.9569	80	0.91854	0.9148
8	0.9346	0.9528	0.00455	0.9528	70	0.89641	0.9100
9	0.9414	0.9533	0.00260	0.9533	68	0.89334	0.9139
10	0.8615	0.9482	0.01993	0.9482	74	0.90526	0.8978
Média	0.9016	0.9496	0.0109	0.9478	72	0.9059	0.9060

1) Análise Discriminante Linear - LDA

[illegible]

2) Perceptron

[illegible]