



Objectivos

- Adquirir a noção de validação da aprendizagem
- Explicar os principais métodos de validação
- Descrever as principais medidas de desempenho
- Adquirir uma noção acerca do processo de realização de experiências em AA



- Sumário:

- Fundamentos de Aprendizagem Automática
 - Avaliação do modelo aprendido
 - Organização de experiências



Validação

- A validação dos algoritmos de aprendizagem constitui um problema importante
- O mesmo consiste em avaliar a proporção do erro que se espera que um algoritmo cometa no caso de um problema determinado



Medidas de desempenho

- No caso de problemas de classificação a medida mais comum é a designada *taxa de erro*
- Se define através da proporção dos exemplos incorrectamente classificados de entre a totalidade dos exemplos de teste



Medidas de desempenho

- Se definem outras medidas que permitem quantificar diferentes tipos de erro
- Considerando um conjunto de teste com N exemplos
 - Para um exemplo positivo
 - Se a predição for positiva \rightarrow verdadeiro positivo (tp)
 - Caso contrario \rightarrow falso negativo (fn)
 - Para um exemplo negativo
 - Se a predição for negativa \rightarrow verdadeiro negativo (tn)
 - Caso contrario \rightarrow falso positivo (fp)



Matriz de confusão

		Classe Verdadeira		
		0 (+)	1 (-)	Total
Classe predita	0 (+)	tp	fp	p'
	1 (-)	fn	tn	n'
Total		p	n	N



Medidas de desempenho

- Taxa de erro $Terr = (fp + fn) / N = e / N$
- Taxa de êxito $Tex = (tp + tn) / N = 1 - Terr$
- Especificidade $Es = tn / (fp + tn) = tn / n$
- Sensibilidade $P_{tp} = tp / (tp + fn) = tp / p$
 - Ou proporção de verdadeiros positivos
- Proporção de falsos positivos $P_{fp} = fp / (fp + tn) = fp / n$



Medidas de desempenho

- No caso de predições numéricas se utiliza outro tipo de medidas
- Mais frequente -> erro quadrático médio

$$E_{qm} = \frac{(p_1 - a_1)^2 + (p_1 - a_2)^2 + \dots + (p_n - a_n)^2}{n}$$



Outros factores a considerar

- Simplicidade do modelo
- Complexidade em termos de tempo e espaço
- Claridade e interpretabilidade da solução
- Facilidade de programação



Métodos de validação

- A taxa de erro determinada a partir dos dados utilizados ao treinar um algoritmo de aprendizagem (conjunto de treino) não é um bom indicador do seu desempenho futuro
- Para avaliar o desempenho futuro é necessário determinar o erro cometido utilizando um conjunto de dados independente que não foi utilizado ao treinar o algoritmo (conjunto de teste)



Métodos de validação

- Em alguns casos a aprendizagem se leva a cabo em duas fases
 - 1ª - definição da estrutura básica do algoritmo
 - 2ª - optimização dos parâmetros envolvidos na estrutura
- Neste caso se utiliza um terceiro conjunto de dados (conjunto de validação) utilizado na 2ª fase da aprendizagem



Holdout

- Consiste na separação dos dados em duas partes, um conjunto de treino e outro de teste
- Geralmente se utiliza $1/3$ dos dados como conjunto de teste e o resto para treinar o algoritmo



Holdout

- Para lidar com factores aleatórios derivados da forma em que os dados são divididos, o método pode ser levado a cabo várias vezes sobre diferentes partições independentes dos dados
- Neste caso a estimação da taxa de erro é feita a partir das médias das taxas obtidas em cada caso



Validação cruzada (*cross validation*)

- Consiste em dividir os dados em K subconjuntos com tamanhos aproximadamente iguais, sendo utilizados $K - 1$ subconjuntos para treinar o algoritmo e o restante para o teste
- A estimação da taxa de erro é feita através da média das taxas obtidas ao repetir o processo K vezes com as diferentes combinações de $K - 1$ subconjuntos de treino
- Geralmente denominada validação cruzada em K folhas (*K-fold cross validation*)

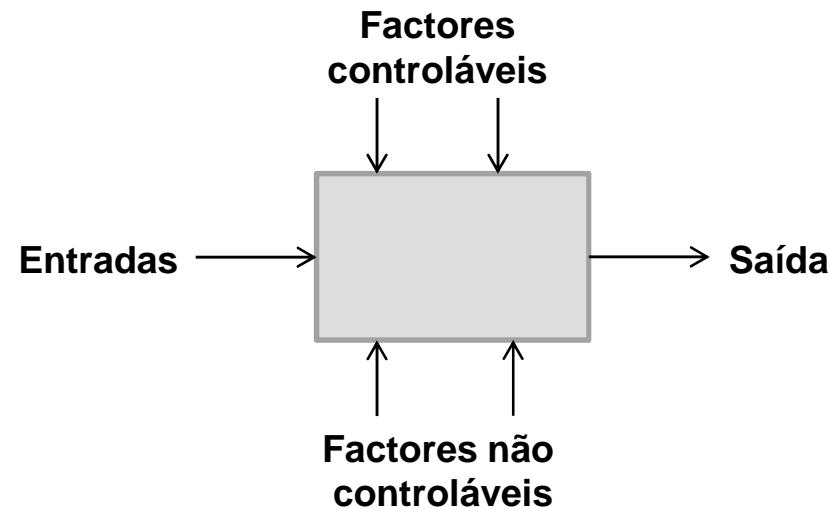


Leave-one-out

- Caso particular da validação cruzada no qual o parâmetro K é igual ao número de exemplos existentes no conjunto de dados
- Neste caso o conjunto de teste é formado por um único exemplo em cada repetição

Experiências em AA

- Em AA geralmente a informação sobre a aplicabilidade de um algoritmo a um problema se obtém de maneira experimental
- Uma experiência é um teste, ou conjunto de testes, no qual são manejados os factores que afectam as saídas e se observam as alterações que ocorrem na resposta com o objectivo de obter informação



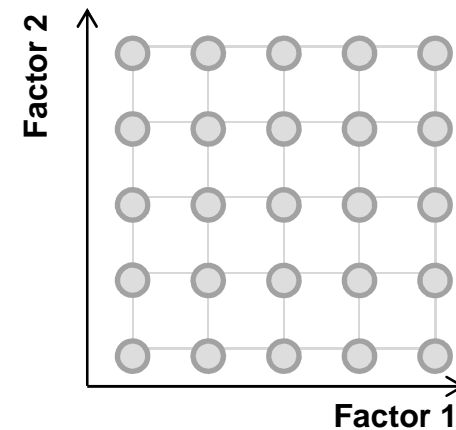
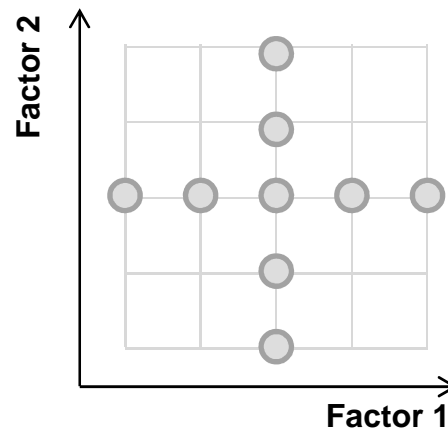
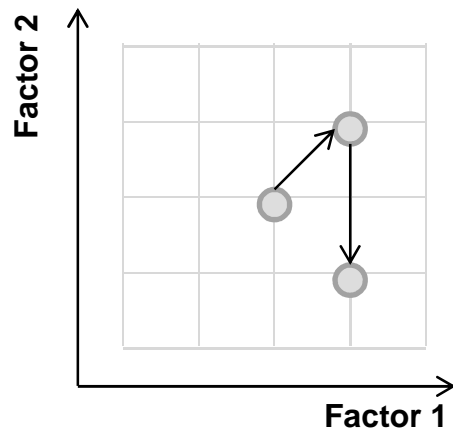


Experiências em AA

- Factores controláveis
 - Algoritmo utilizado, hiperparâmetros do algoritmo, conjunto de treino, características de entrada...
- Factores não controláveis
 - Ruído existente nos dados, subconjunto de treino utilizado em particular, inicialização dos algoritmos...
- Variável de resposta
 - Gerada a partir da saída
 - Erro médio de classificação sobre um conjunto de teste...

Estratégias de experimentação

- Melhor tentativa (*best guess*)
- Um factor de cada vez
- Desenho factorial





Directrizes para a realização de experiências

- Determinar o objectivo do estudo
- Seleccionar a variável de resposta
- Escolha dos factores e níveis
- Escolha do desenho experimental
- Realização da experiência
- Análise estatística dos dados
- Conclusões e recomendações



Bibliografia

- Witten
 - Validação, pgs. 147 – 150; 152 – 157; 180 - 182
- Alpaydin
 - Desenho experiências, pgs. 475 – 481; 483 – 486