

ICONE – EPUSP: Grupo de Inteligência Computacional,
Modelagem e Neurocomputação Eletrônica

Site do Grupo ICONE --- <http://www.lsi.usp.br/ICONE/>
... e facebook: [//www.facebook.com/ICONE.EPUSP/](https://www.facebook.com/ICONE.EPUSP/)

Meu e-mail: emilio.delmoral@usp.br / emilio@lsi.usp.br



5

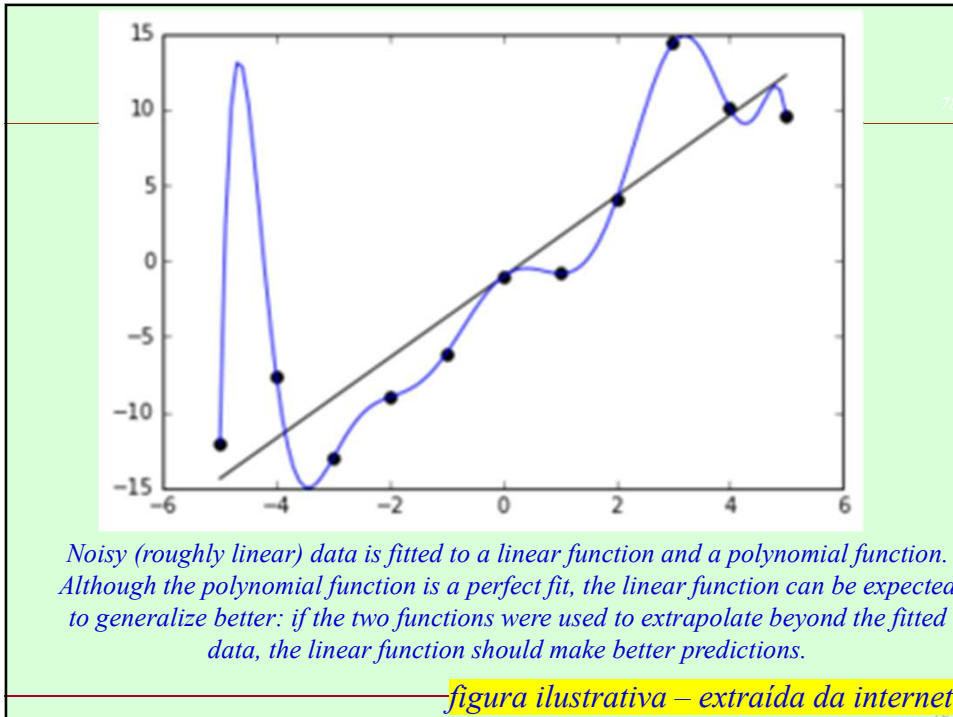
Um tema importante:

Sobreapredizado / Sobreajuste

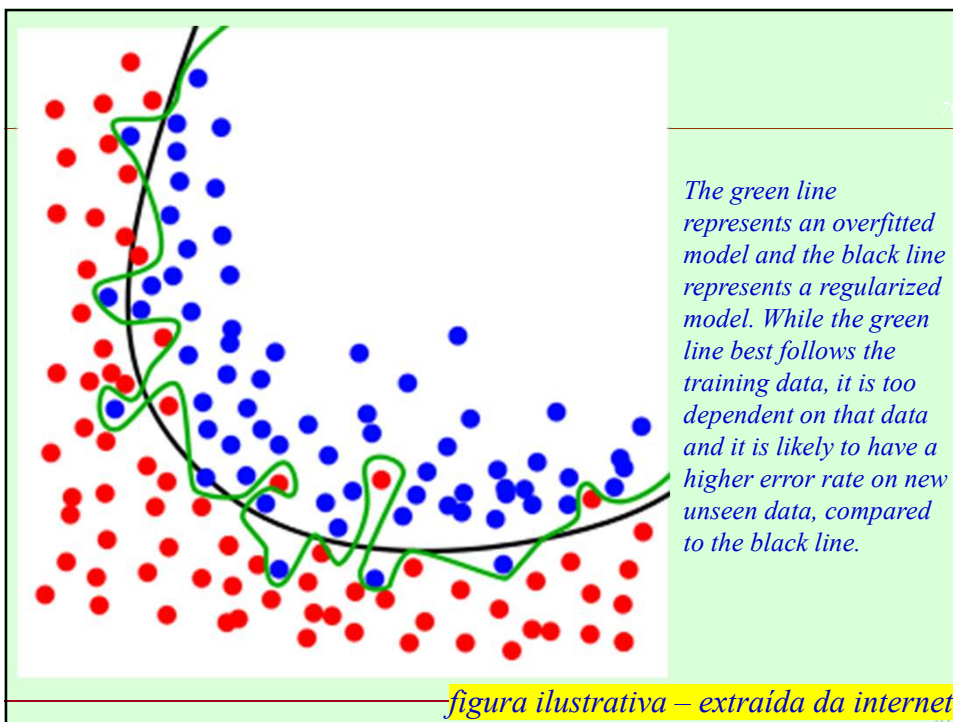
***Conceito, entendimento da sua
origem e formas de limitá-lo***

© Prof. Emilio Del Moral – EPUSP

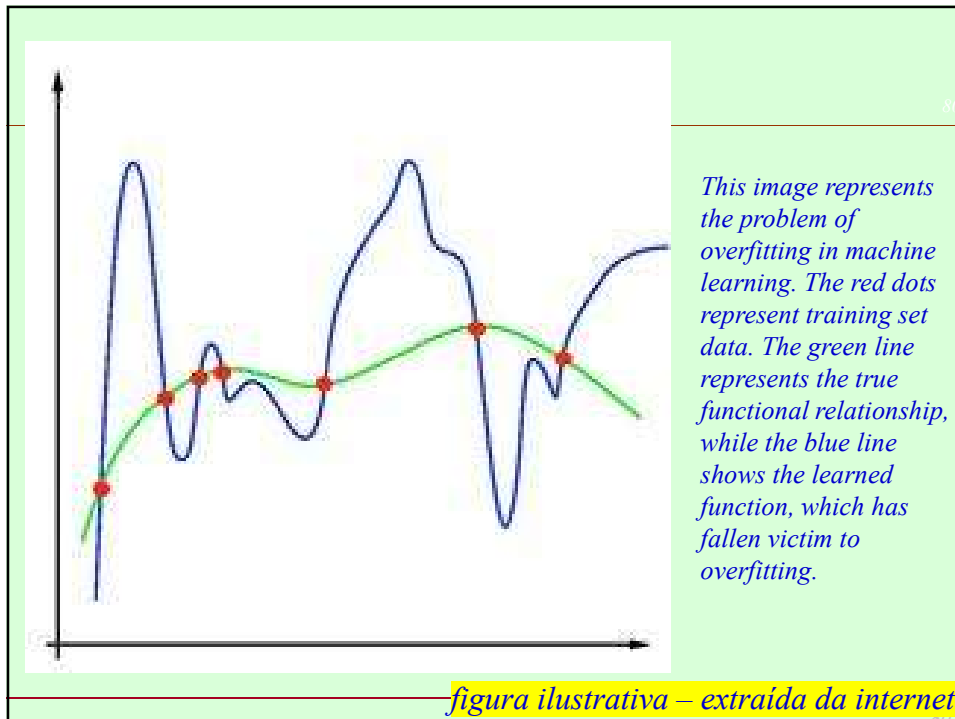
77



78



79



80

Sobreaprendizado em polinômios:

Primeiro entendamos o conceito num universo mais familiar (e mais simples), o de regressão polinomial univariada, usada para representar dados com comportamento linear ou não linear e sujeitos a alguma flutuação em “y”

...

Depois, vocês mesmos podem pensar nos equivalentes dos nossos raciocínios feitos aqui para o universo de polinômios, mas para o universo de RNAs e mesmo de outros tipos de modelos com número de parâmetros variável (complexidade variável) que você conheça ...

81
© Prof. Emilio Del Moral – EPUSP

81

Falemos em lousa um pouco sobre a reta média para um conjunto de pares (x,y), a parábola média, a cúbica média ... etc

$$y \sim ax+b ; \quad y \sim ax^2 +bx +c ; \quad y \sim ax^3 +bx^2 +cx +d$$

e mais além, falemos sobre regressão polinomial univariada, com o grau do polinômio aproximador podendo ser 1, 2, 3, ou mesmo graus bastante mais altos como 50, 51 etc.

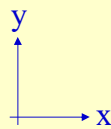
$$y \sim ax^{51} +bx^{50} +cx^{49} + \dots$$

82

© Prof. Emilio Del Moral – EPUSP

82

Sobreaprendizado ilustrado em sua fenomenologia, na regressão polinomial (de vários graus) univariada



Os dados empíricos (x^i, y^i) estão em verde;

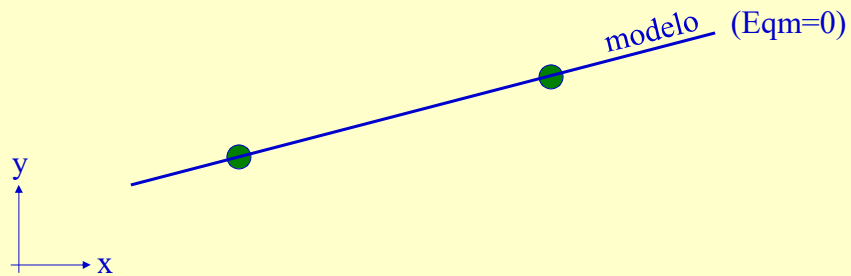
83

© Prof. Emilio Del Moral – EPUSP

83

Sobreaprendizado ilustrado em sua fenomenologia, na regressão polinomial (de vários graus) univariada

façamos uso de modelagem linear ...



Os dados empíricos (x^i, y^i) estão em verde;
O modelo linear gerado a partir dos dados, em azul.

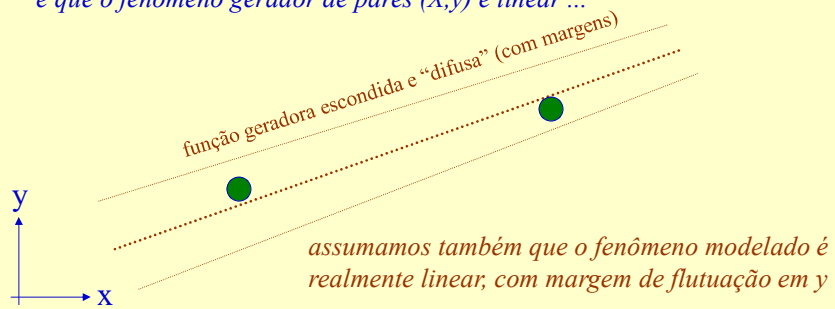
84

© Prof. Emilio Del Moral – EPUSP

84

Sobreaprendizado ilustrado em sua fenomenologia, na regressão polinomial (de vários graus) univariada

e que o fenômeno gerador de pares (X,y) é linear ...



Os dados empíricos (x^i, y^i) estão em verde;
O modelo linear gerado a partir dos dados, em azul.
O fenômeno gerador de pares (x,y) é linear em essência, mas tem alguma flutuação randômica em y . A tendência e os limites da flutuação estão representados em marrom

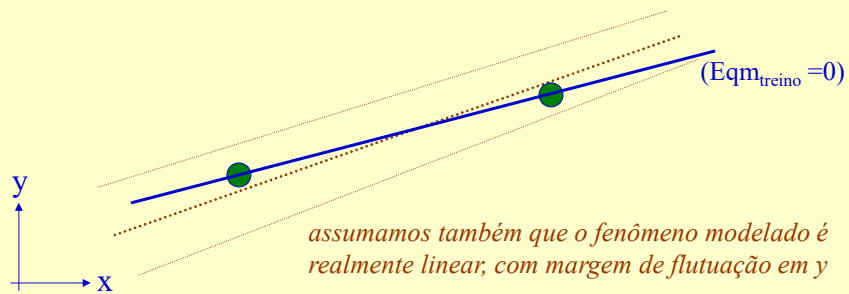
85

© Prof. Emilio Del Moral – EPUSP

85

Sobreaprendizado ilustrado em sua fenomenologia, na regressão polinomial (de vários graus) univariada

façamos uso de modelagem linear ...



Os dados empíricos (x^i, y^i) estão em verde;
O modelo linear gerado a partir dos dados, em azul.
O fenômeno gerador de pares (x, y) é linear em essência, mas tem alguma flutuação randômica em y . A tendência e os limites da flutuação estão representados em marrom

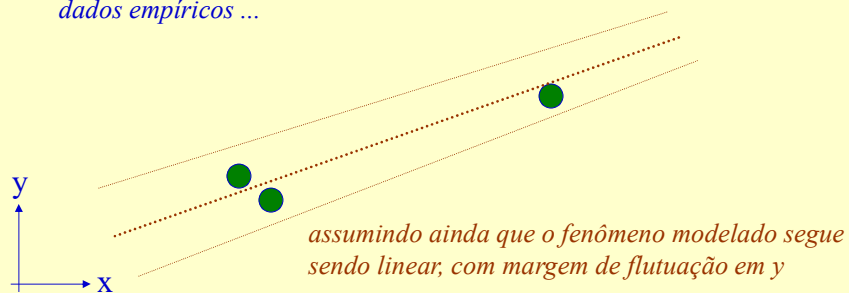
86

© Prof. Emilio Del Moral – EPUSP

86

Sobreaprendizado ilustrado em sua fenomenologia, na regressão polinomial (de vários graus) univariada

Consideremos agora nova situação com mais dados empíricos ...



Os dados empíricos (x^i, y^i) estão em verde;

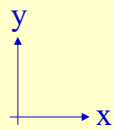
87

© Prof. Emilio Del Moral – EPUSP

87

Sobreaprendizado ilustrado em sua fenomenologia, na regressão polinomial (de vários graus) univariada

Consideremos agora nova situação com mais dados empíricos ... mas na modelagem não se sabe se é o fenômeno linear ou quadrático ...



Os dados empíricos (x^i, y^i) estão em verde;

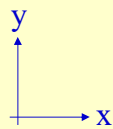
88

© Prof. Emilio Del Moral – EPUSP

88

Sobreaprendizado ilustrado em sua fenomenologia, na regressão polinomial (de vários graus) univariada

Consideremos agora nova situação com mais dados empíricos ... mas na modelagem não se sabe se é o fenômeno linear ou quadrático ...



Os dados empíricos (x^i, y^i) estão em verde;

$(Eqm_{treino} > 0)$

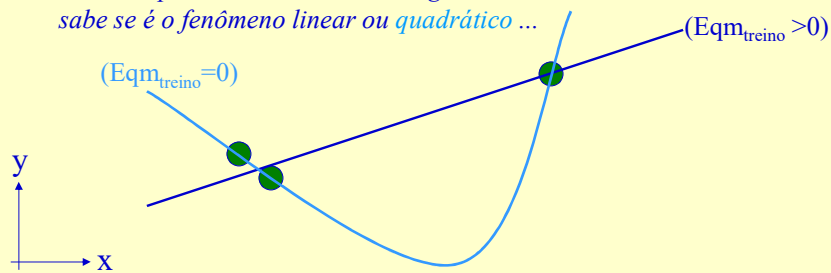
89

© Prof. Emilio Del Moral – EPUSP

89

Sobreaprendizado ilustrado em sua fenomenologia, na regressão polinomial (de vários graus) univariada

Consideremos agora nova situação com mais dados empíricos ... mas na modelagem não se sabe se é o fenômeno linear ou quadrático ...



Os dados empíricos (x^i, y^i) estão em verde;
Dois modelos polinomiais gerados a partir dos dados, em azuis.
O fenômeno gerador de pares (x, y) é linear em essência, mas tem alguma flutuação randômica em y . A tendência e os limites da flutuação estão representados em marrom

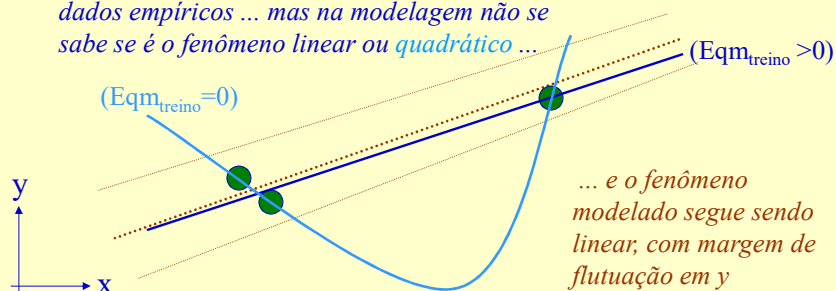
90

© Prof. Emilio Del Moral – EPUSP

90

Sobreaprendizado ilustrado em sua fenomenologia, na regressão polinomial (de vários graus) univariada

Consideremos agora nova situação com mais dados empíricos ... mas na modelagem não se sabe se é o fenômeno linear ou quadrático ...



... e o fenômeno modelado segue sendo linear, com margem de flutuação em y

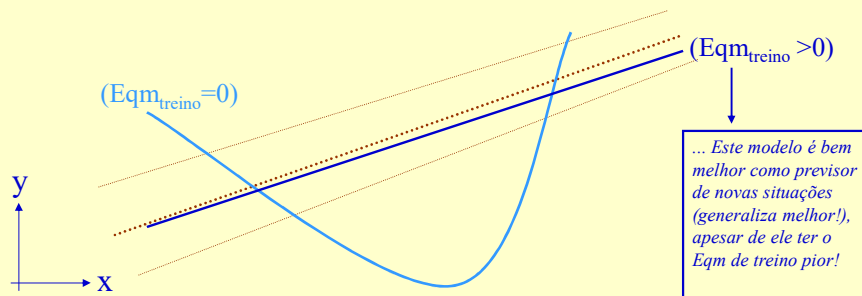
Os dados empíricos (x^i, y^i) estão em verde;
Dois modelos polinomiais gerados a partir dos dados, em azuis.
O fenômeno gerador de pares (x, y) é linear em essência, mas tem alguma flutuação randômica em y . A tendência e os limites da flutuação estão representados em marrom

91

© Prof. Emilio Del Moral – EPUSP

91

Sobreaprendizado ilustrado em sua fenomenologia, na regressão polinomial (de vários graus) univariada



Os dados empíricos (x^i, y^i) estão em verde;
Dois modelos polinomiais gerados a partir dos dados, em azuis.
O fenômeno gerador de pares (x, y) é linear em essência, mas tem alguma flutuação randômica em y . A tendência e os limites da flutuação estão representados em marrom

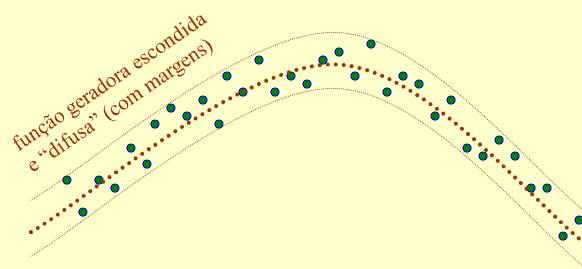
92

© Prof. Emilio Del Moral – EPUSP

92

Sobreaprendizado ilustrado em sua fenomenologia, na regressão polinomial (de vários graus) univariada

um novo exemplo



Os dados empíricos (x^i, y^i) estão em verde;
O fenômeno gerador de pares (x, y) é quadrático em essência, mas tem alguma flutuação randômica em y . A tendência e os limites da flutuação estão representados em marrom

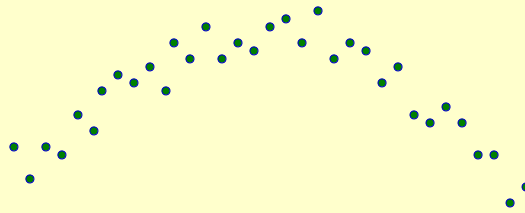
93

© Prof. Emilio Del Moral – EPUSP

93

Sobreaprendizado ilustrado em sua fenomenologia, na regressão polinomial (de vários graus) univariada

um novo exemplo



Os dados empíricos (x^i, y^i) estão em verde;

O fenômeno gerador de pares (x, y) é quadrático em essência, mas tem alguma flutuação randômica em y . A tendência e os limites da flutuação estão representados em marrom

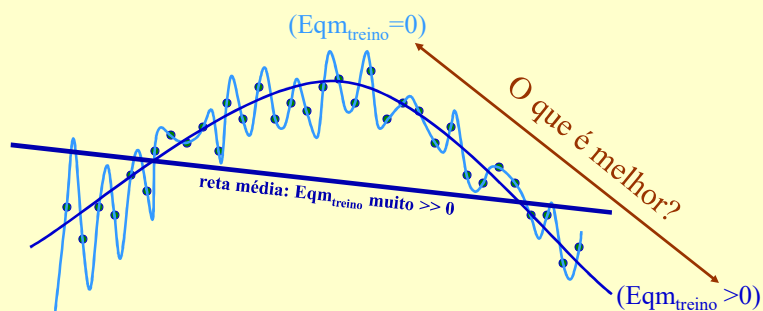
94

© Prof. Emilio Del Moral – EPUSP

94

Sobreaprendizado ilustrado em sua fenomenologia, na regressão polinomial (de vários graus) univariada

um novo exemplo



Os dados empíricos (x^i, y^i) estão em verde;

Três modelos polinomiais gerados a partir dos dados, em azuis.

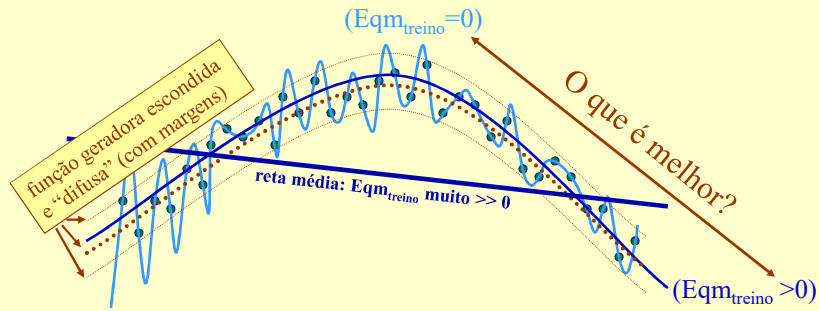
95

© Prof. Emilio Del Moral – EPUSP

95

Sobreaprendizado ilustrado em sua fenomenologia, na regressão polinomial (de vários graus) univariada

um novo exemplo



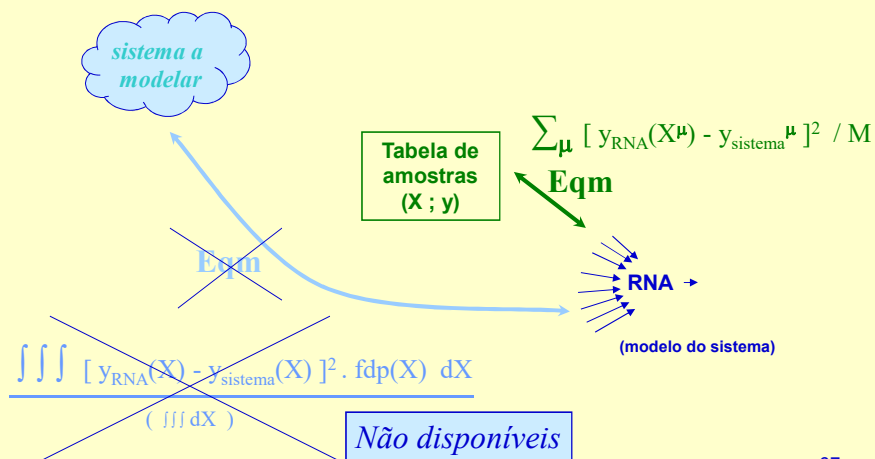
Os dados empíricos (x^μ, y^μ) estão em verde;
Três modelos polinomiais gerados a partir dos dados, em azuis.

96

© Prof. Emilio Del Moral – EPUSP

96

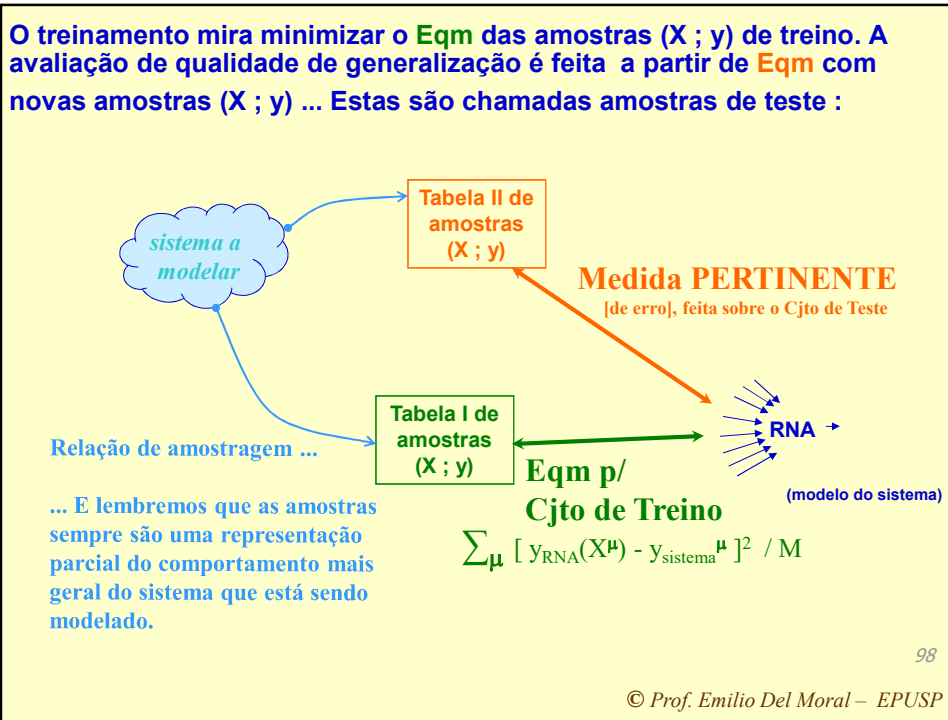
Sistema ... Amostras de treino ... RNA ...



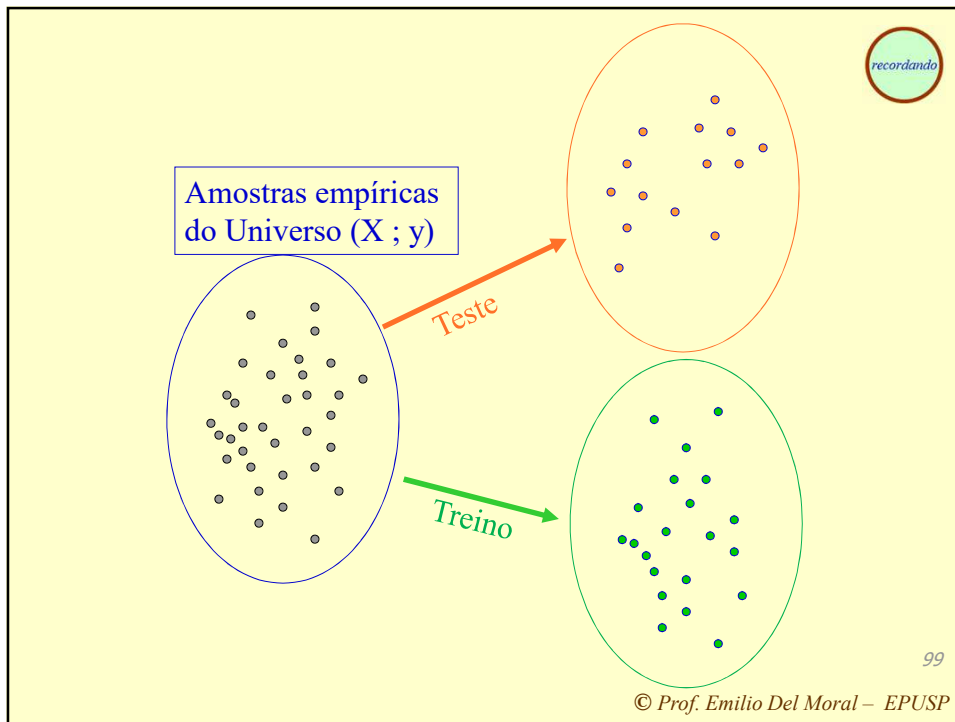
97

© Prof. Emilio Del Moral Hernandez

97

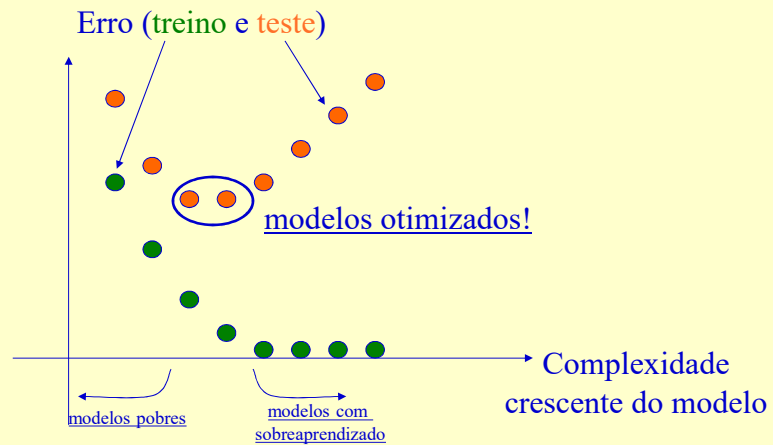


98



99

Sobreaprendizado em “sumário executivo”



100

© Prof. Emilio Del Moral – EPUSP

100

Revisitando os Conjuntos de Dados Empíricos ...

Treino + Teste ...

versus

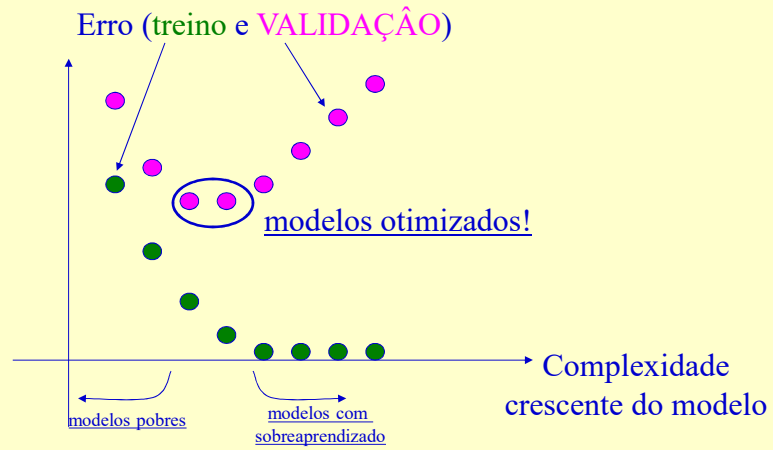
Treino + Validação + Teste

101

© Prof. Emilio Del Moral – EPUSP

101

Sobreaprendizado em “sumário executivo”

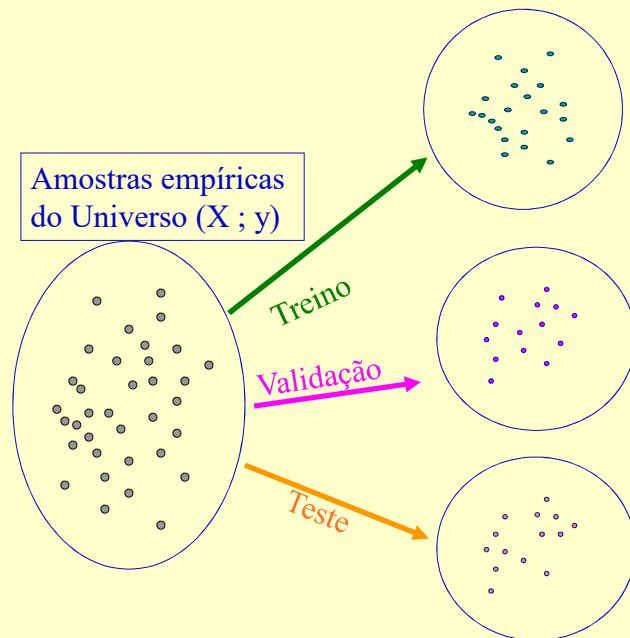


102

© Prof. Emilio Del Moral – EPUSP

102

Amostras empíricas do Universo ($X ; y$)

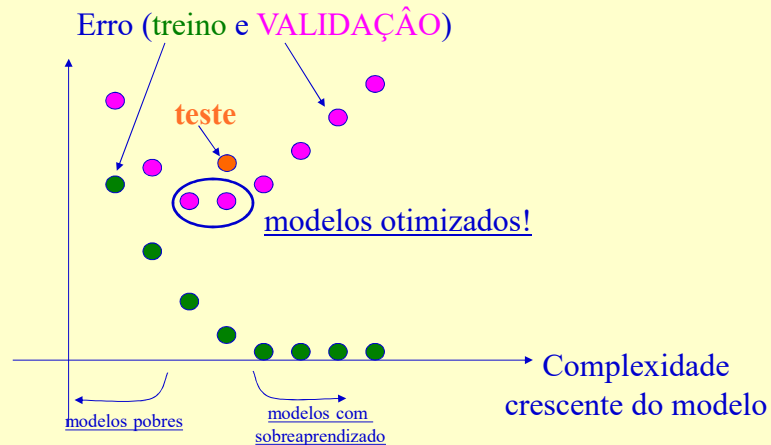


103

© Prof. Emilio Del Moral – EPUSP

103

Sobreaprendizado em “sumário executivo”



104

© Prof. Emilio Del Moral – EPUSP

104

Alguns dos diversos usos do conceito de conjunto de validação, adicional ao conjunto de treino e de teste:

- *Seleção de complexidade do modelo neural para limitação de sobreaprendizado*
- *Ativação do early stop no aprendizado (Matlab): critério de parada adicional no processo de refinamento de pesos sinápticos*
- *Balizador no processo de seleção de estratégias de pré-processamento / extração de medidas X alternativos*
- *... etc*

105

© Prof. Emilio Del Moral – EPUSP

105

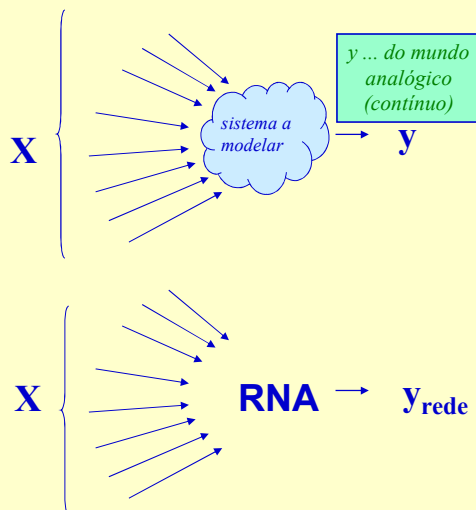
Identificando os ingredientes para o risco de sobreaprendizado nos contextos de regressão multivariada e de reconhecimento de padrões multivariado

107

© Prof. Emilio Del Moral – EPUSP

107

Modelagem de um sistema por função de mapeamento $X \rightarrow y$ (a RNA como regressor analógico não linear multivariável)



Assumimos que a variável y do sistema a modelar é uma função (normalmente desconhecida e possivelmente não linear) de diversas outras variáveis desse mesmo sistema

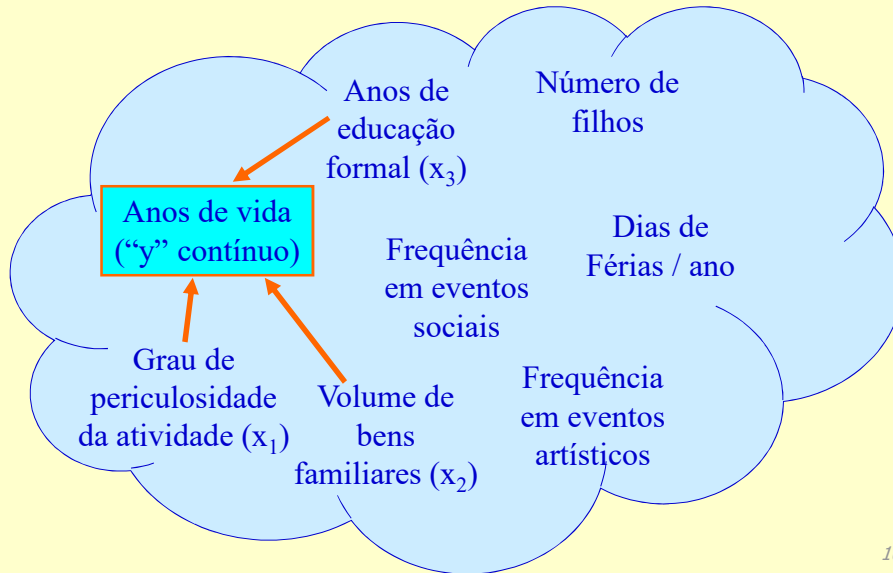
A RNA, para ser um bom modelo do sistema, deve reproduzir essa relação entre X e y , tão bem quanto possível

108

© Prof. Emilio Del Moral – EPUSP

108

Um hipotético universo de variáveis interdependentes, passível de modelagem/ens

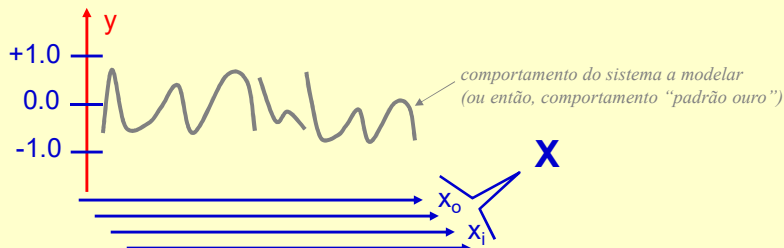


109

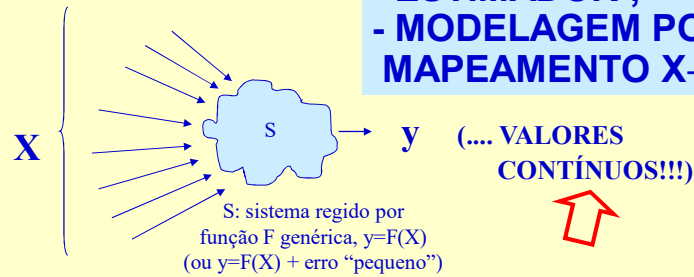
© Prof. Emilio Del Moral – EPUSP

109

A função $y(X)$ “a descobrir”, num caso geral de função analógica $y(X)$



**- ESTIMADOR ;
- MODELAGEM POR
MAPEAMENTO $X \rightarrow y$**

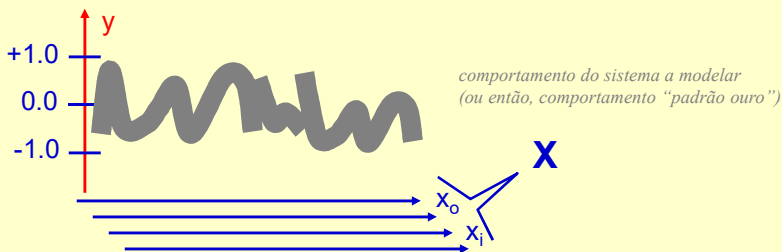


110

© Prof. Emilio Del Moral – EPUSP

110

Cenário mais real: a “função” $y(X)$ do sistema modelado é “difusa”: $y = F_{\text{médio}}(X) + \text{flutuação} \dots$



.... em problemas concretos / reais, há sempre alguma ambiguidade no mapeamento que leva valores de X a valores de y . Para decepção de Cybenko, não temos uma função $y = F(X)$ no sentido matemático exato, pois para uma dada ênupla de valores X fixados, temos tipicamente uma faixa de valores que podem ser observados para a variável y : $y = F_{\text{médio}}(X) + \text{flutuação}$.

Neste cenário, buscamos que o modelo capture o comportamento médio das relações observadas entre X e y :

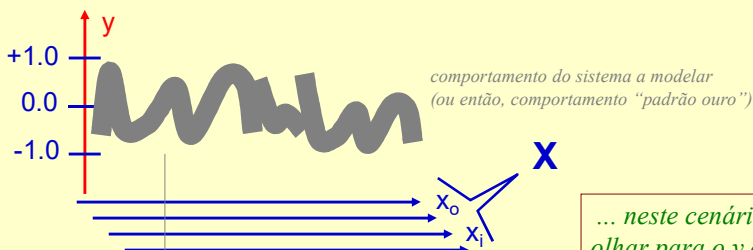
$$\dots y_{\text{rede}} \sim y_{\text{médio}} \text{ esperado para um dado } X$$

111

© Prof. Emilio Del Moral – EPUSP

111

Cenário mais real: a “função” $y(X)$ do sistema modelado é “difusa”: $y = F_{\text{médio}}(X) + \text{flutuação} \dots$



... neste cenário, podemos olhar para o y observado no sistema que se deseja modelar não mais como um valor específico bem definido, mas como um valor médio esperado (dado valor de X) e uma faixa de valores em torno desse valor médio esperado.

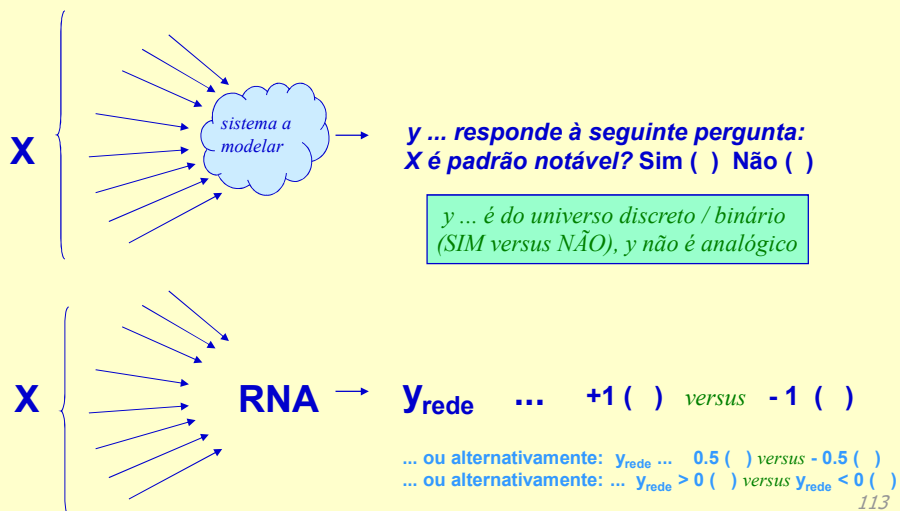
112

© Prof. Emilio Del Moral – EPUSP

112

RNAs como reconhecedor / detetor de padrões

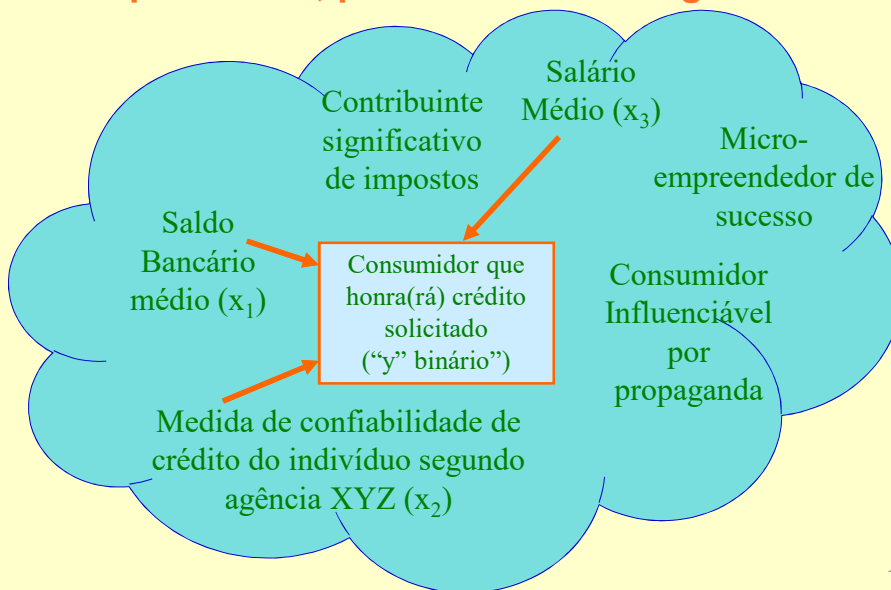
...



© Prof. Emilio Del Moral – EPUSP

113

Um hipotético universo de variáveis interdependentes, passível de modelagem/ens

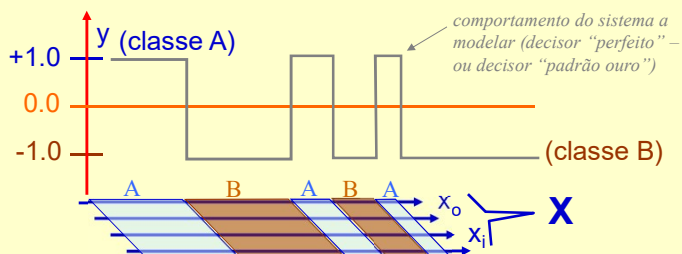


114

© Prof. Emilio Del Moral – EPUSP

114

Caso de classificação binária / reconhecimento de padrões, será do tipo ...



**CLASSIFICADOR;
RECONHECEDOR DE
PADRÕES NOTÁVEIS**



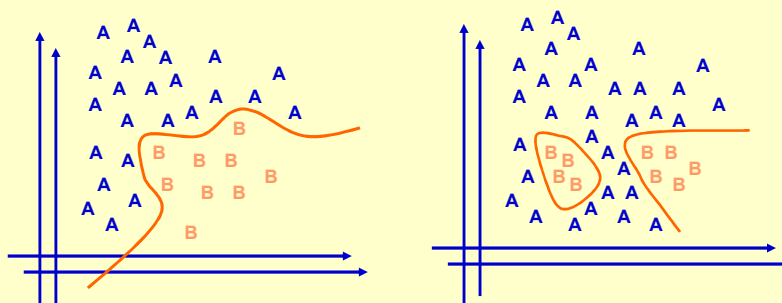
115

© Prof. Emilio Del Moral – EPUSP

115

Capacidade de reconhecimento de padrões em casos complexos NÃO LINEARES

Com as RNAs, a hipersuperfície de separação entre classes vai muito além dos hiperplanos

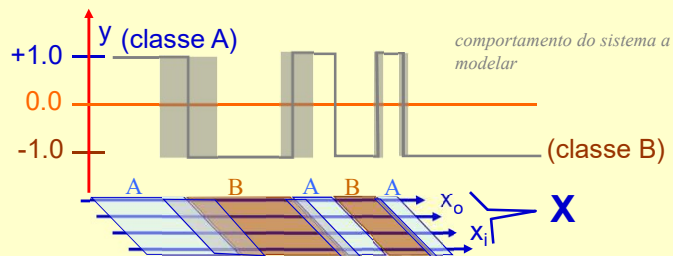


116

© Prof. Emilio Del Moral – EPUSP

116

Cenário mais real: a separação entre regiões do espaço de X não é perfeitamente definida



.... em problemas concretos / reais, há sempre alguma ambiguidade no mapeamento que leva valores de X aos valores discretos de y . Não temos uma função $y=F(X)$ no sentido matemático exato, pois para uma dada ênupla de valores X fixado temos em alguns casos de fronteira a possibilidade de observar no y empírico tanto a classe A quanto a classe B: $y=A$ ou B, com maior ou menor probabilidade para cada classe de acordo com o X . Neste desejamos que o modelo capture o comportamento médio das relações observadas entre X e y :

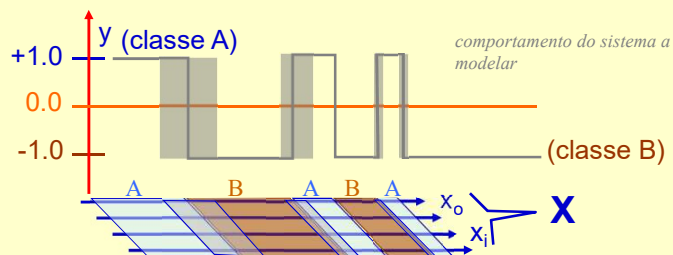
... $y_{rede} \sim$ classe 'mais esperada' para um dado X

117

© Prof. Emilio Del Moral – EPUSP

117

Cenário mais real: a separação entre regiões do espaço de X não é perfeitamente definida



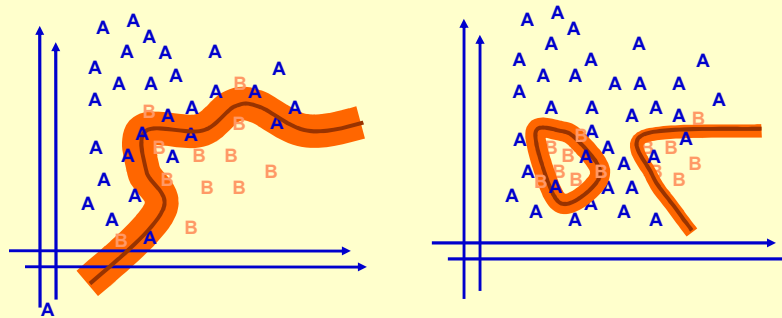
... podemos olhar para o y (classe A ou B) observado no sistema que se deseja modelar não mais como uma classe sempre bem definida e com fronteiras de separação entre A e B bem definidas no espaço de valores de X , mas como sendo delineadas na modelagem através de fronteiras com eventuais faixas de tolerância e com sobreposição parcial das classes no espaço de X

118

© Prof. Emilio Del Moral – EPUSP

118

Situações de classes com sobreposição parcial no espaço de atributos X ; situações de fronteiras de separação difusas ...



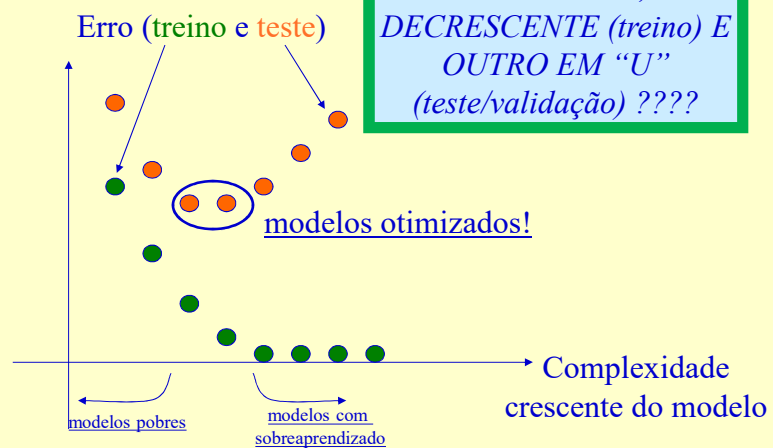
120

© Prof. Emilio Del Moral – EPUSP

120

Sobreaprendizado em “sumário executivo”

É FÁCIL OBTER ESSES DOIS GRÁFICOS, UM DECRESCENTE (treino) E OUTRO EM “U” (teste/validação) ????



121

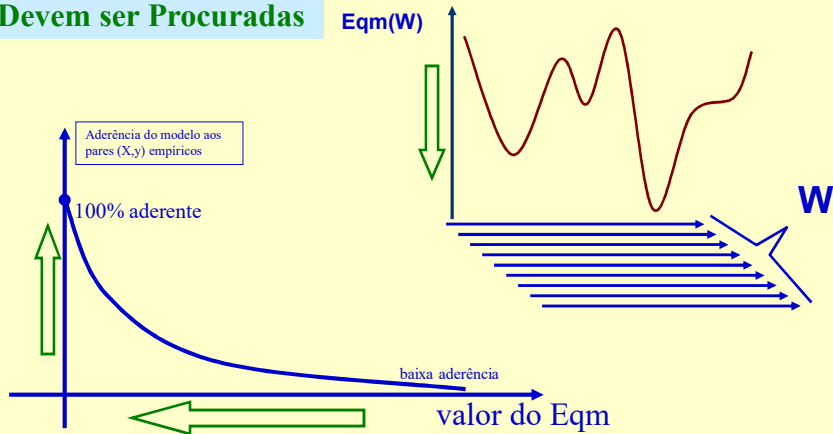
© Prof. Emilio Del Moral – EPUSP

121

O que devemos mirar quando exploramos o espaço de pesos W buscando que a RNA seja um bom modelo?

Devemos buscar Maximização da aderência = Mínimo Eqm possível

**As Setas Verdes
Indicam Situações que
Devem ser Procuradas**

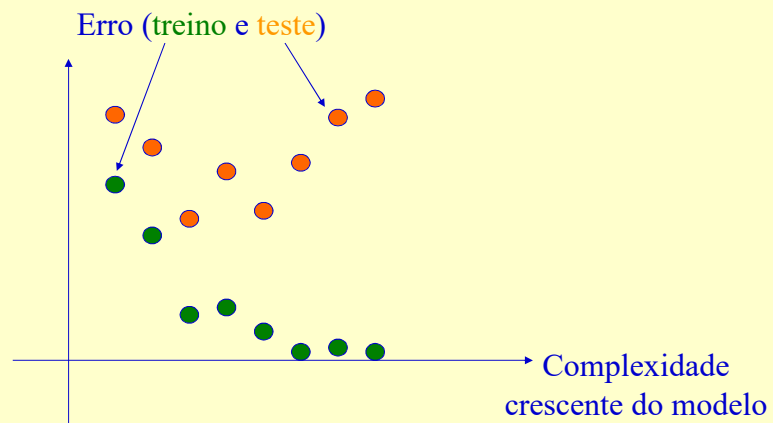


124

© Prof. Emilio Del Moral – EPUSP

124

Atenção para componentes randômicas que impactam muito quando se faz um único ensaio de medida de erro, para cada tamanho de rede específico (um ensaio apenas, para cada grau de complexidade) ...



125

© Prof. Emilio Del Moral – EPUSP

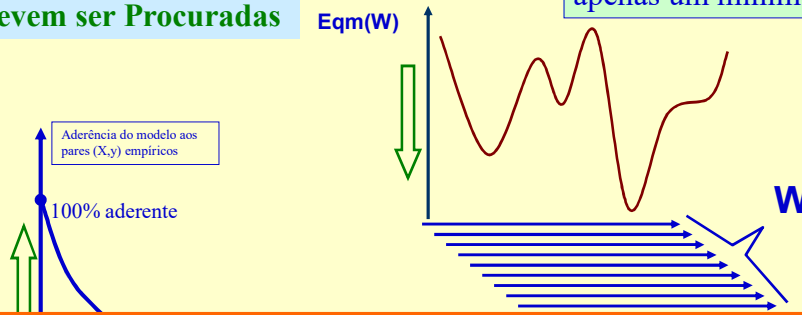
125

O que devemos mirar quando exploramos o espaço de pesos W buscando que a RNA seja um bom modelo?

Devemos buscar Maximização da aderência = Mínimo Eqm possível

**As Setas Verdes
Indicam Situações que
Devem ser Procuradas**

Será que temos apenas um mínimo??



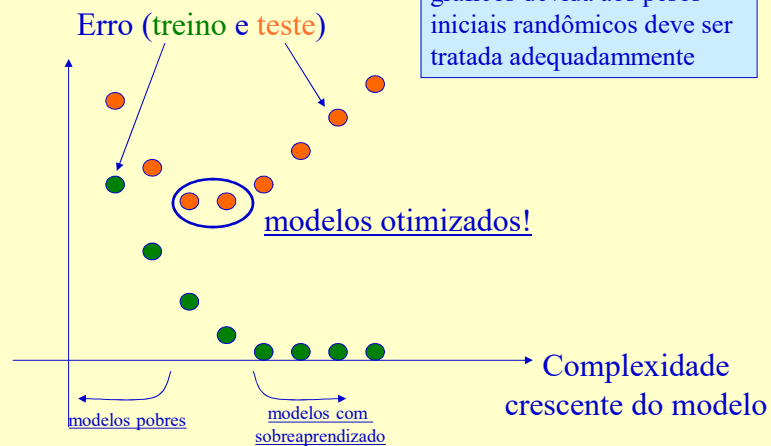
... Para não sermos reféns de mínimos locais com alto Eqm, podemos aplicar o gradiente descendente repetidamente na mesma RNA, com novos pesos iniciais randômicos em cada rodada, mantendo para o modelo final apenas os valores de pesos associados ao ensaio com o melhor dos resultados finais no Eqm!

© Prof. Emilio Del Moral – EPUSP

126

Com repetidos ensaios em cada grau de complexidade os mínimos locais são evitados e detectamos adequadamente o sobreaprendizado

Ou seja, a flutuação nos gráficos devida aos pesos iniciais randômicos deve ser tratada adequadamente



127

© Prof. Emilio Del Moral – EPUSP

127