

Intervalo de 10 mins ...

*Após o intervalo – Entendimento
detalhado do aprendizado de MLPs
(calibração de seus pesos w 's) por EBP
– Error Back Propagation!*

*(para quem já fez PSI3471 comigo –
disciplina de 4º ano –, é recordação e
opcional, mas pode ser divertido!)*

*Aprendizado em RNAs do
tipo MLP – Multi Layer
Perceptron – através do
algoritmo Error Back
Propagation*

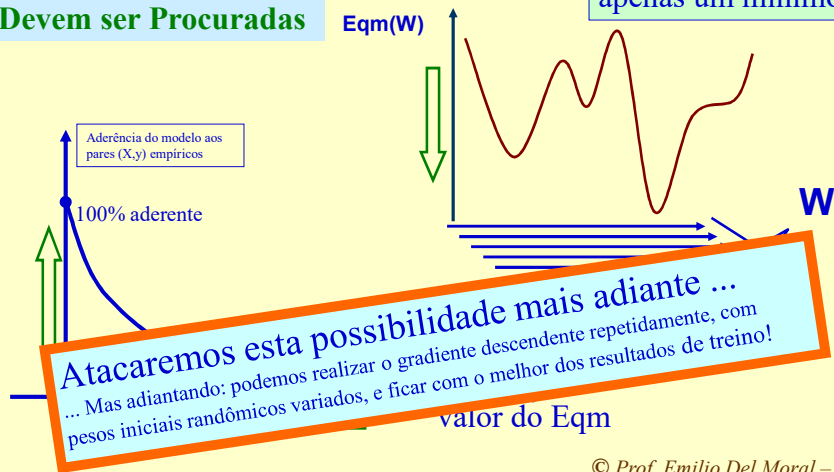
Como escolhemos os valores dos diversos w 's ?

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

Devemos mirar Maximização da aderência = Mínimo E_{qm} possível

As Setas Verdes Indicam Situações que Devem ser Procuradas

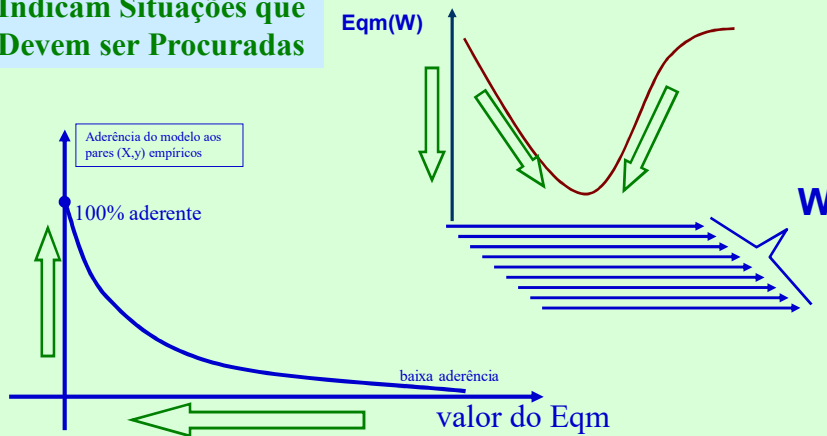
Será que temos apenas um mínimo??



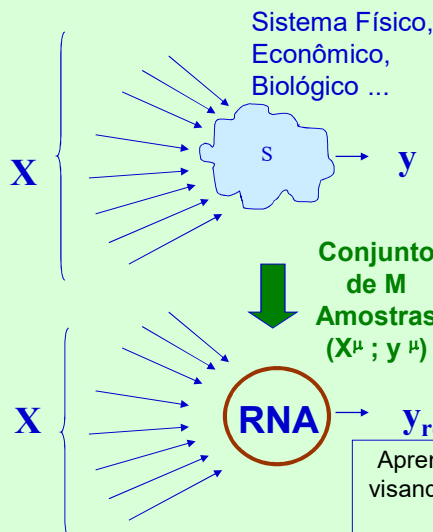
O que devemos buscar quando exploramos o espaço de pesos W buscando que a RNA seja um bom modelo?

Devemos buscar Maximização da aderência = Mínimo Eqm possível

As Setas Verdes Indicam Situações que Devem ser Procuradas



Conjunto de treino em arquiteturas supervisionadas (ex. clássico: MLP com Error Back Propagation)



A computação desejada da rede pode ser definida simplesmente através de amostras / exemplos do comportamento requerido

$$Eqm = \frac{1}{M} \sum_{\mu=1}^M (y_{rede}(\vec{X}^{\mu}, \vec{W}) - y^{\mu})^2$$

$$\vec{\Delta W} = -\eta \cdot \vec{\nabla} Eqm$$

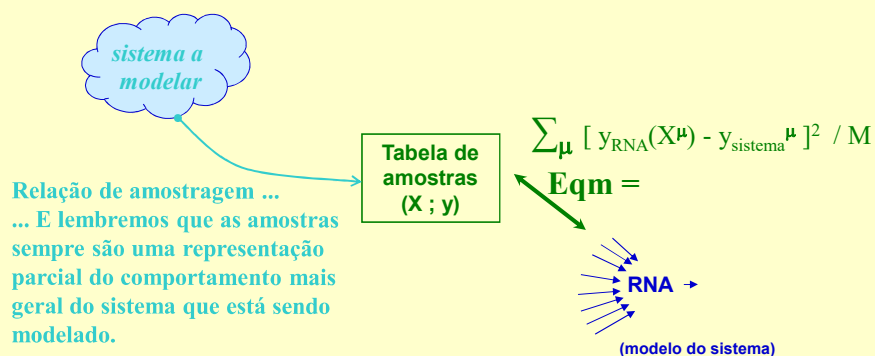
... em loop ...

Aprendizado: Espaço de pesos W é explorado visando aproximar ao máximo a computação da rede da computação desejada

Deduzindo as Equações do Aprendizado em RNAs do tipo MLP – Multi Layer Perceptron – com o algoritmo Error Back Propagation (Gradiente Descendente)

© Prof. Emilio Del Moral – EPUSP

O treinamento mira minimizar o **Eqm** das amostras (X ; y) de treino. (exclusivamente!)



12

© Prof. Emilio Del Moral – EPUSP

Um Exemplo Ilustrativo para o Conceito de Conjunto de Treinamento e dos M pares (X,y) ...

13

© Prof. Emilio Del Moral – EPUSP

Exemplo de regressão multivariada para estimação contínua usando MLP

- O valor do y contínuo ... neste exemplo corresponde ao volume de consumo futuro num dado tipo de produto "A" a ser ofertado pela empresa a um cliente corrente já consumidor de outros produtos da empresa ("B" e "C"), volume esse previsto com base em várias medidas quantitativas que caracterizam tal indivíduo. ... Assim, $y = \text{Consumo do Produto A} = F(x_1, x_2, x_3, x_4, x_5)$.
- Consideremos 4 variáveis de entrada no modelo preditivo neural, ou seja, temos 5 medidas em X :
 - x_1 : Idade do indivíduo
 - x_2 : Renda mensal do indivíduo
 - x_3 : Volume de clicks do indivíduo no website de exibição de produtos oferecidos pela empresa
 - x_4 : Volume de consumo desse cliente observado para outro Produto B da mesma empresa
 - x_5 : Volume de consumo desse cliente Produto C da mesma empresa
- Problema: desenvolver uma MLP para regressão contínua multivariada que permita estimar esse volume de consumo futuro y com base no conhecimento dos X e numa base de dados de aprendizado com esses dados X e y para 350 já clientes de universo populacional similar ao do novo consumidor potencial. 14

© Prof. Emilio Del Moral – EPUSP

Exemplo de dados empíricos tabulados em Excel ...

Cliente (μ)	Idade (x_1)	Renda (x_2)	Clics (x_3)	Consumo do Produto B (x_4)	Consumo do Produto C (x_5)	Consumo do Produto A (y)
1	50	78	302	958	136	9800
2	65	128	186	985	196	8760
3	57	150	221	1093	35	520
....
M-2	16	19	51	707	131	11640
M-1	30	75	7	29	78	9640
M	19	47	116	285	124	5320

15

© Prof. Emilio Del Moral – EPUSP

Equivalente em .txt, em formato apropriado para o ambiente Multiple Back Propagation ...

Cliente (μ)	Idade (x_1)	Renda (x_2)	Clics (x_3)	Consumo do Produto B (x_4)	Consumo do Produto C (x_5)	Consumo do Produto A (y)
1	50	78	302	958	136	9800
2	65	128	186	985	196	8760
3	57	150	221	1093	35	520
....
M-2	16	19	51	707	131	11640
M-1	30	75	7	29	78	9640
M	19	47	116	285	124	5320

*Equivalente em txt
Para uso do MBP*

```

treino em txt para exemplo de consumo A e B - Bloco de notas
Arquivo Editar Formatar Exibir Ajuda
Idade Renda Clics ConsumoA ConsumoB ConsumoA
50 78 302 958 136 9800
65 128 186 985 196 8760
57 150 221 1093 35 520
(...)
16 19 51 707 131 11640
30 75 7 29 78 9640
19 47 116 285 124 5320
    
```

75
Moral – EPUSP

*A estratégia de Aprendizado para o MLP
mais conhecida:*

Error Back Propagation (EBP)

= Propagação Reversa de Erro

*= Método do Gradiente personalizado
ao Eqm(W) do MLP*

17

© Prof. Emilio Del Moral – EPUSP

*Mas entendamos PRIMEIRO
o que é o método numérico do
gradiente ascendente /
gradiente descendente
genérico,*

*que pode ser aplicado tanto para se
chegar paulatinamente ao máximo de
uma função quanto para se chegar ao
mínimo de uma função
(ascendente / descendente)*

18

© Prof. Emilio Del Moral – EPUSP

Chamada oral sobre a lição de casa: estudar / reestudar os conceitos e a parte operacional de derivadas parciais, do vetor Gradiente, e da regra da cadeia ...

- Derivadas parciais (que são as componentes do gradiente):

$$\frac{\partial f(a,b,c)}{\partial a} \quad \frac{\partial f(a,b,c)}{\partial b} \quad \frac{\partial f(a,b,c)}{\partial c}$$

- Vetor Gradiente, útil ao método do máximo declive:

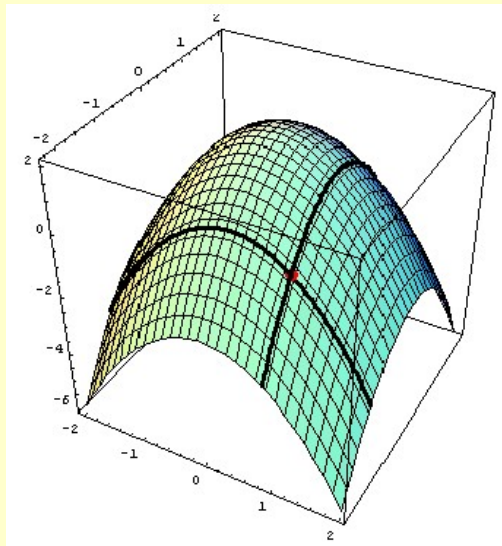
$$\left(\frac{\partial \text{Eqm}(W)}{\partial w_1}, \frac{\partial \text{Eqm}(W)}{\partial w_2}, \frac{\partial \text{Eqm}(W)}{\partial w_3}, \dots \right) \quad \Delta \vec{W} = -\eta \cdot \vec{\nabla} \text{Eqm}_-$$

- Regra da cadeia, necessária ao cálculo de derivadas quando há encadeamento de funções:

$$\frac{\partial f(g(h(a)))}{\partial a} = \frac{\partial f}{\partial g} \cdot \frac{\partial g}{\partial h} \cdot \frac{\partial h}{\partial a}$$

© Prof. Emilio Del Moral – EPUSP

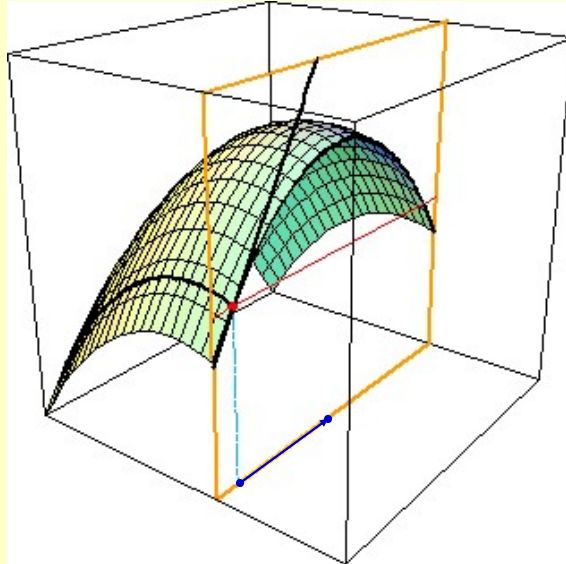
Derivada parcial- ilustração p/ função de 2 variáveis apenas



A visual model of the partial derivative

© Prof. Emilio Del Moral – EPUSP

Derivada parcial- ilustração p/ função de 2 variáveis apenas

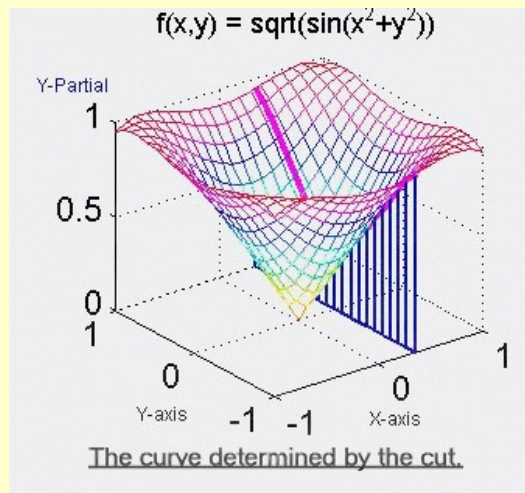


A visual model of the partial derivative

21

© Prof. Emilio Del Moral – EPUSP

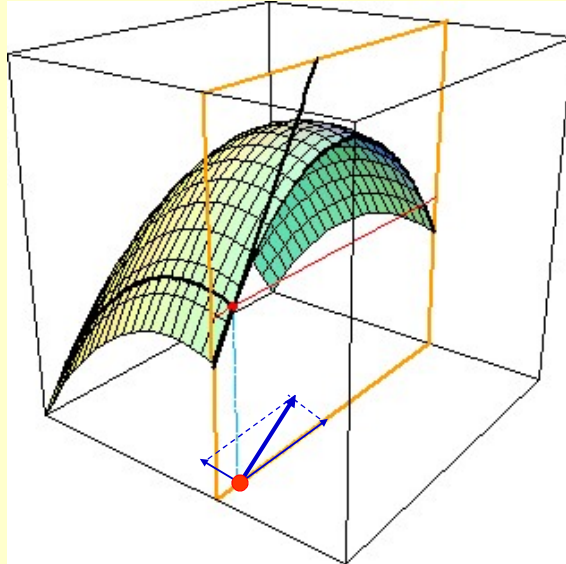
mais ilustrações p/ a derivada parcial em função de 2 variáveis



A visual model of the partial derivative with respect to y. 22

© Prof. Emilio Del Moral – EPUSP

Formação do vetor gradiente a partir de duas derivadas parciais



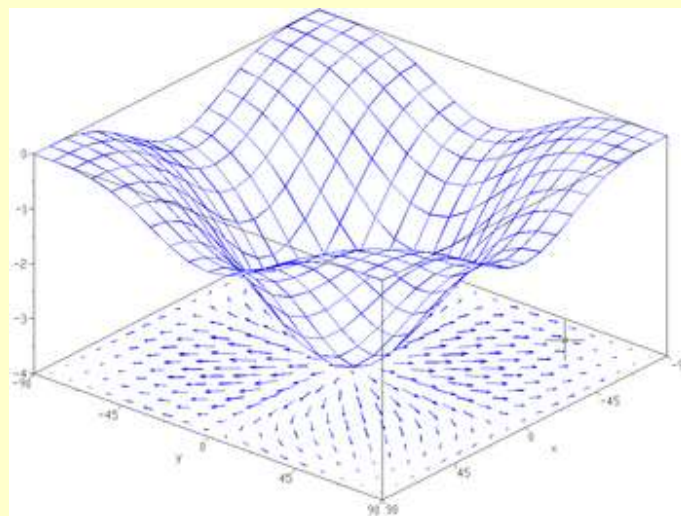
A visual model of the partial derivative

23

© Prof. Emilio Del Moral – EPUSP

<http://en.wikipedia.org/wiki/Gradient>

... O vetor gradiente indica a direção ascendente e seu módulo a magnitude de crescimento da função escalar – ilustração p/ função de 2 variáveis apenas



24

© Prof. Emilio Del Moral – EPUSP

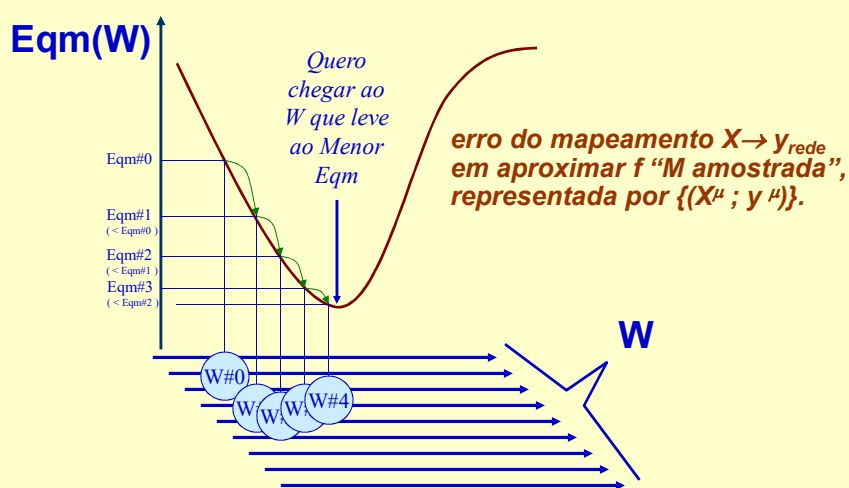
Processo de refinamentos graduais a cada iteração ...

W#0	Eqm#0	GradEqm(W#0)	DeltaW#0 = - n.GradEqm(W#0)
W#1 (= W#0 + DeltaW#0)	Eqm#1 (< Eqm#0)	GradEqm(W#1)	DeltaW#1 = - n.GradEqm(W#1)
W#2 (= W#1 + DeltaW#1)	Eqm#2 (< Eqm#1)	GradEqm(W#2)	DeltaW#2 = - n.GradEqm(W#2)
W#3 (= W#2 + DeltaW#2)	Eqm#3 (< Eqm#2)	GradEqm(W#3)	DeltaW#3 = - n.GradEqm(W#3)
W#4 (= W#3 + DeltaW#3)	Eqm#4 (< Eqm#3)	GradEqm(W#4)	DeltaW#4 = - n.GradEqm(W#4)
...
W#k (= W#k-1 + DeltaW#k-1)	Eqm#k (< Eqm#k-1)	GradEqm(W#k)	DeltaW#k = - n.GradEqm(W#k)
...

25

© Prof. Emilio Del Moral – EPUSP

A estratégia do EBP / Gradiente Descendente no aprendizado do MLP



26

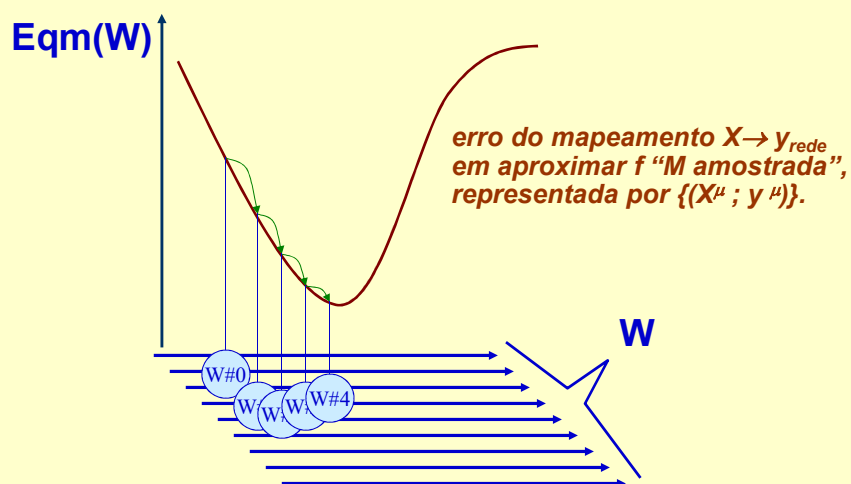
© Prof. Emilio Del Moral – EPUSP

Este gráfico é denso e toca em muitos aspectos interrelacionados ... revisitemos alguns desses aspectos isoladamente com focos específicos nessas revisitas, assim teremos gráficos algo mais simples de interpretar ...

27

© Prof. Emilio Del Moral – EPUSP

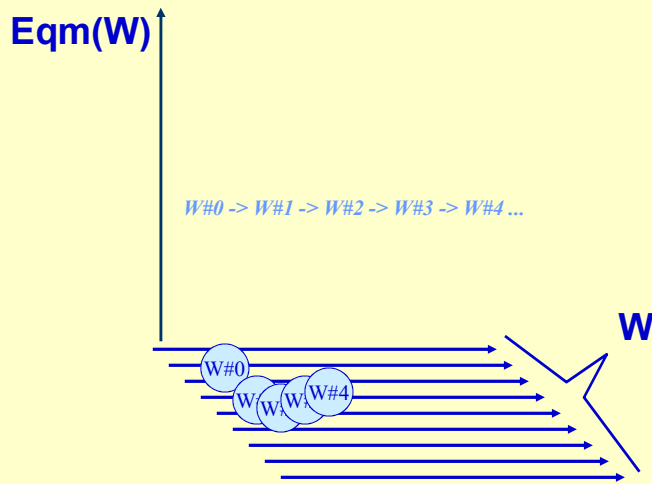
A estratégia do EBP / Gradiente Descendente no aprendizado do MLP



28

© Prof. Emilio Del Moral – EPUSP

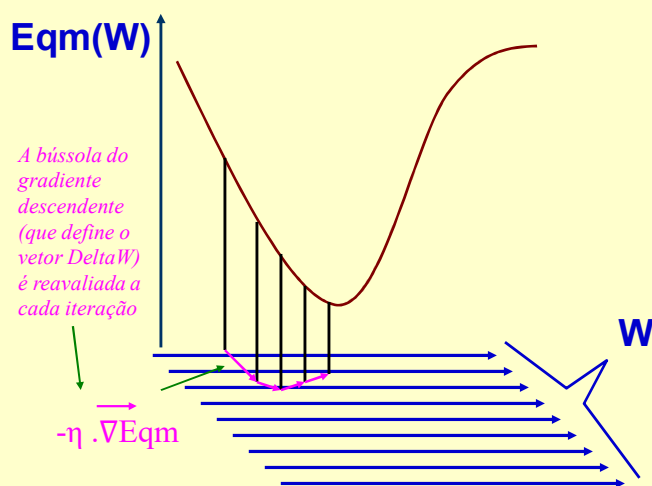
Foco na evolução dos w 's com as iterações ...



29

© Prof. Emilio Del Moral – EPUSP

Foco nos diferentes Δw de cada iteração ...



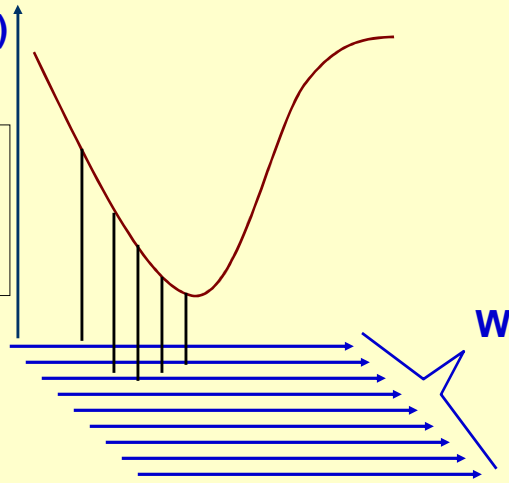
30

© Prof. Emilio Del Moral – EPUSP

Foco na evolução do Eqm com as iterações ...

Eqm(W)

Por construção, o Eqm decresce a cada iteração



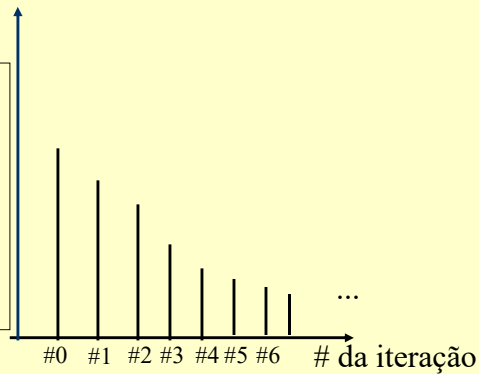
31

© Prof. Emilio Del Moral – EPUSP

Plotando a evolução do Eqm com as iterações ...

Eqm(#)

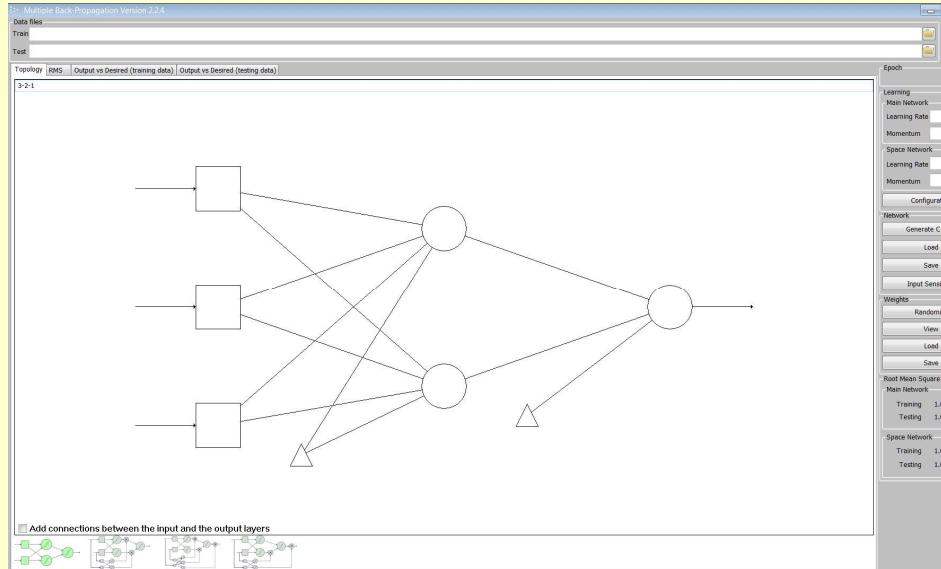
Por construção, Eqm decresce a cada iteração, até estabilização em ponto de mínimo



32

© Prof. Emilio Del Moral – EPUSP

Como o que estamos aprendendo no domínio de equações analíticas se relaciona com o exercitado no ambiente MBP?



© Prof. Emilio Del Moral – EPUSP

Gráfico fornecido pelo ambiente MBP da evolução do Eqm com o número de repetidos usos da “bússola do gradiente descendente”:
isto conecta o MBP com o gráfico apresentado no slide anterior

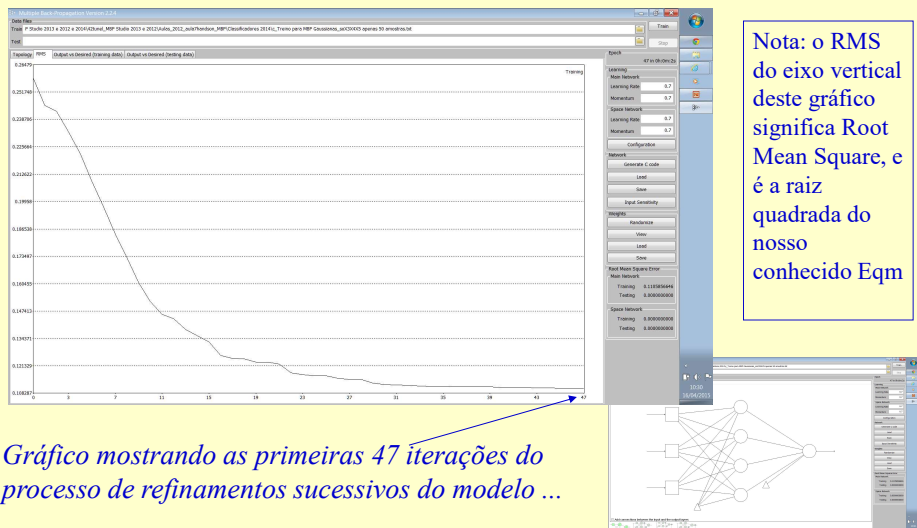
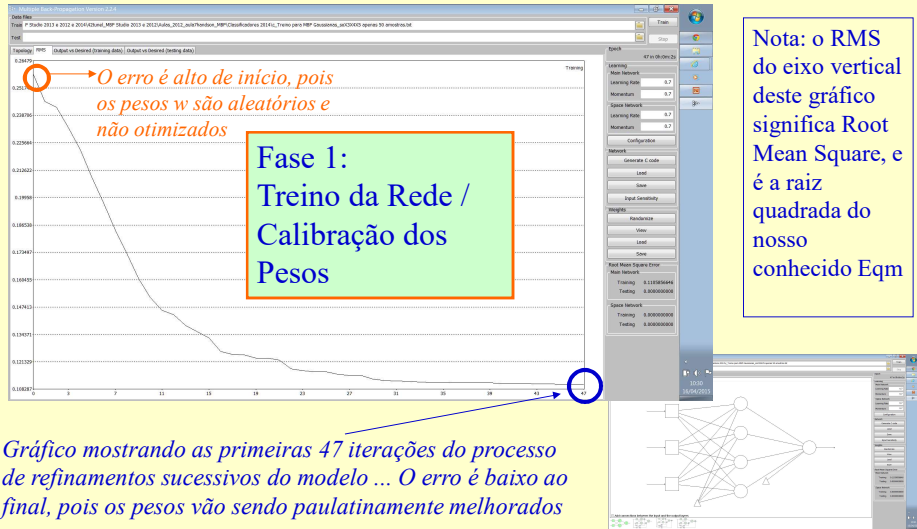


Gráfico mostrando as primeiras 47 iterações do processo de refinamentos sucessivos do modelo ...

Nota: o RMS do eixo vertical deste gráfico significa Root Mean Square, e é a raiz quadrada do nosso conhecido Eqm

© Prof. Emilio Del Moral – EPUSP

Gráfico fornecido pelo ambiente MBP da evolução do Eqm com o número de repetidos usos da “bússola do gradiente descendente”:
isto conecta o MBP com o gráfico apresentado no slide anterior



Nota: o RMS do eixo vertical deste gráfico significa Root Mean Square, e é a raiz quadrada do nosso conhecido Eqm

Gráfico mostrando as primeiras 47 iterações do processo de refinamentos sucessivos do modelo ... O erro é baixo ao final, pois os pesos vão sendo paulatinamente melhorados

© Prof. Emilio Del Moral – EPUSP

Fim do extra – A aula

5

de PSI3571 em
02-abr-2018

**terminou aqui
para todos alunos**

36

Reconhecimento de Padrões, Modelagem, Inteligência Computacional – Prof. Emilio Del Moral H. – ©

O Ciclo completo da modelagem:

0) Formalização do problema, mapeamento quantitativo em um modelo neural inicial e ... 0b) coleta de pares empíricos (X,y)

1) Fase de TREINO da RNA (MLP): com conhecimento dos X e dos y, que são ambos usados na calibração do modelo

2) Fase de TESTE / Caracterização da qualidade da RNA para generalizar: temos novos pares X e y, com y guardado “na gaveta”, usado apenas para avaliação, não para re-calibração. É como um ensaio de uso final do modelo, com possibilidade de medir a sua qualidade com o y que foi guardado na gaveta.

[Fase de refinamentos da RNA, dados e modelo, em ciclos, desde 0]

3) Fase de USO FINAL da RNA, com y efetivamente não conhecido, e estimado com conhecimento dos X + uso do modelo calibrado.

.... Diferenças e semelhanças entre 1, 2 e 3

37

© Prof. Emilio Del Moral – EPUSP