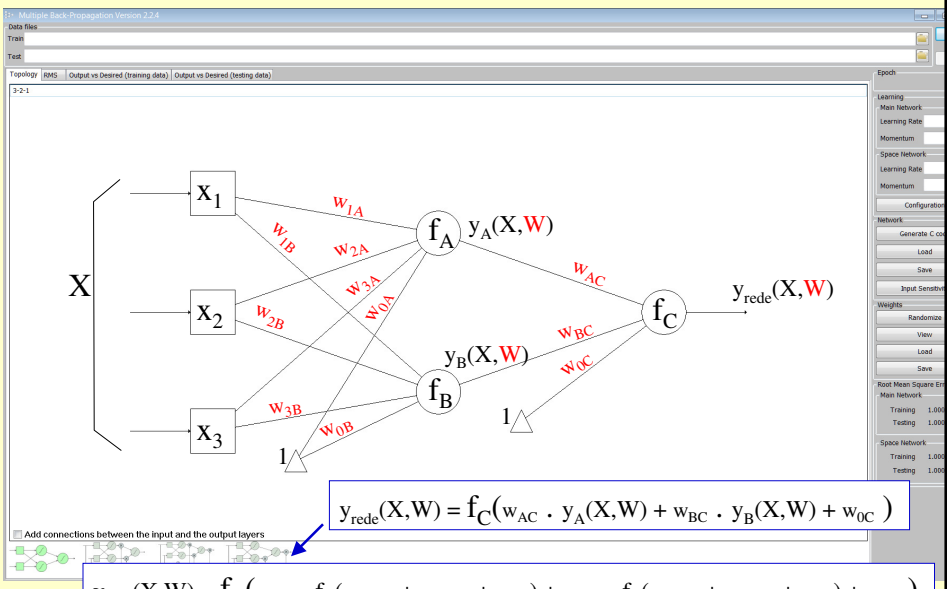
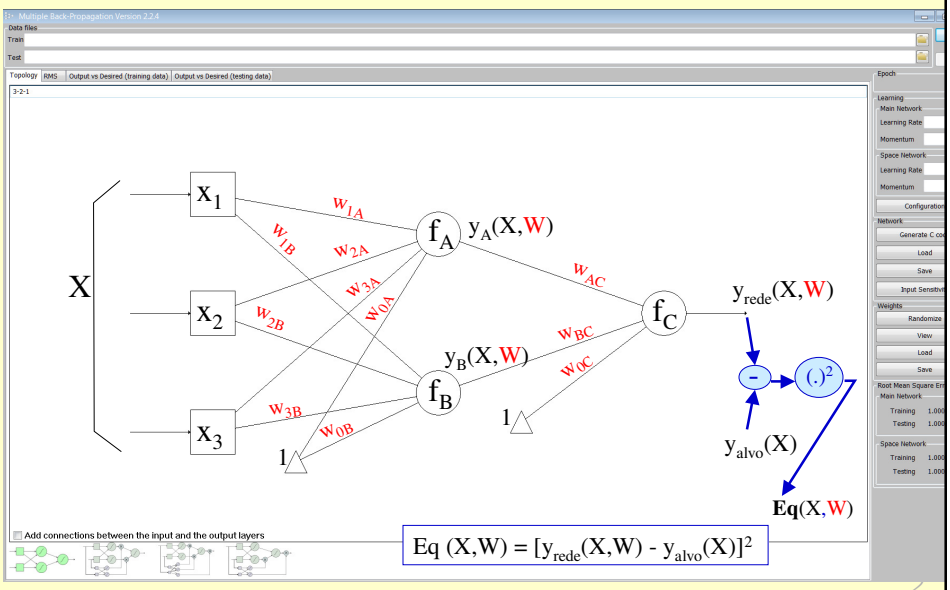


$$y_{rede}(X,W) = f_C\left(w_{AC} \cdot y_A(X,W) + w_{BC} \cdot y_B(X,W) + w_{0C}\right)$$

$$y_{rede}(X,W) = f_C\left(w_{AC} \cdot f_A(w_{1A}.x_1 + w_{2A}.x_2 + w_{0A}) + w_{BC} \cdot f_B(w_{1B}.x_1 + w_{2B}.x_2 + w_{0B}) + w_{0C}\right)$$



$$Eq\ (X,W) = [y_{rede}(X,W) - y_{alvo}(X)]^2$$

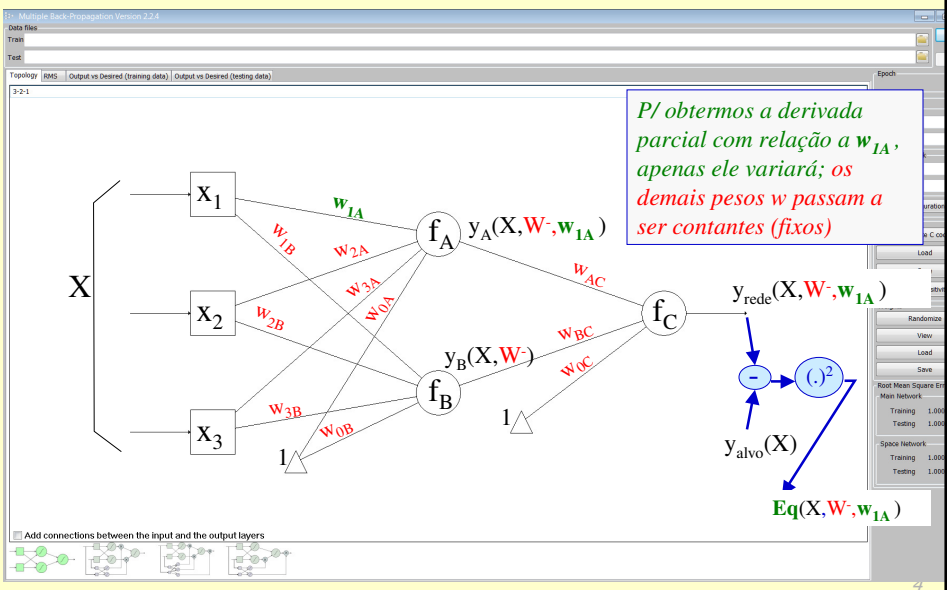
Invertamos o operador gradiente e a somatória

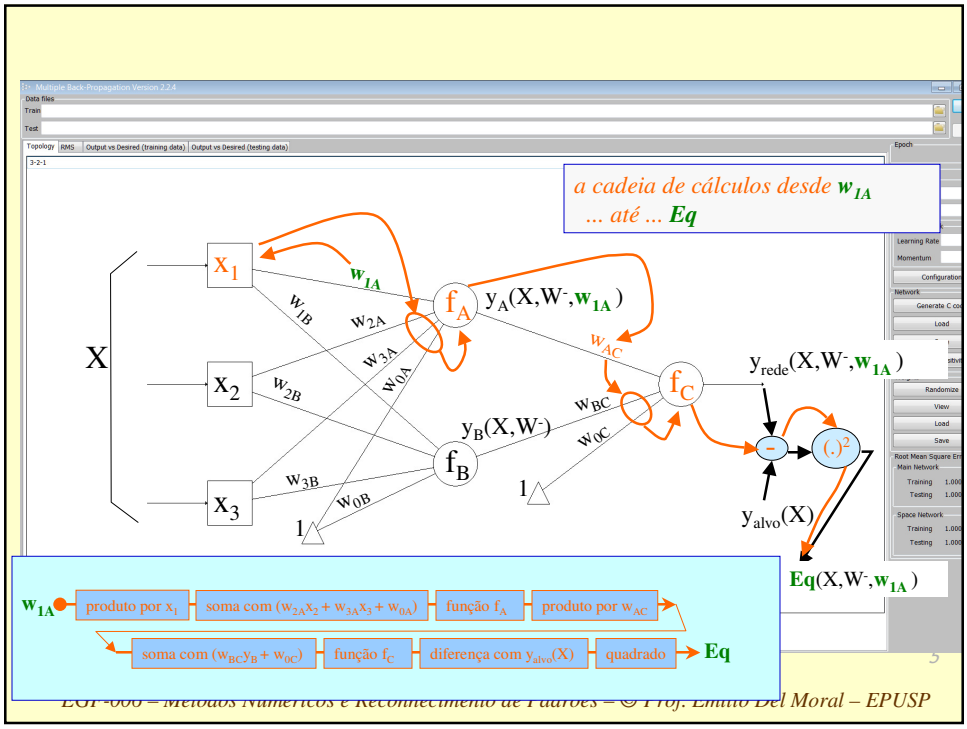
... afinal, gradiente é uma derivada, e a derivada de um soma de várias funções é igual à soma das derivadas individuais de cada componente da soma:

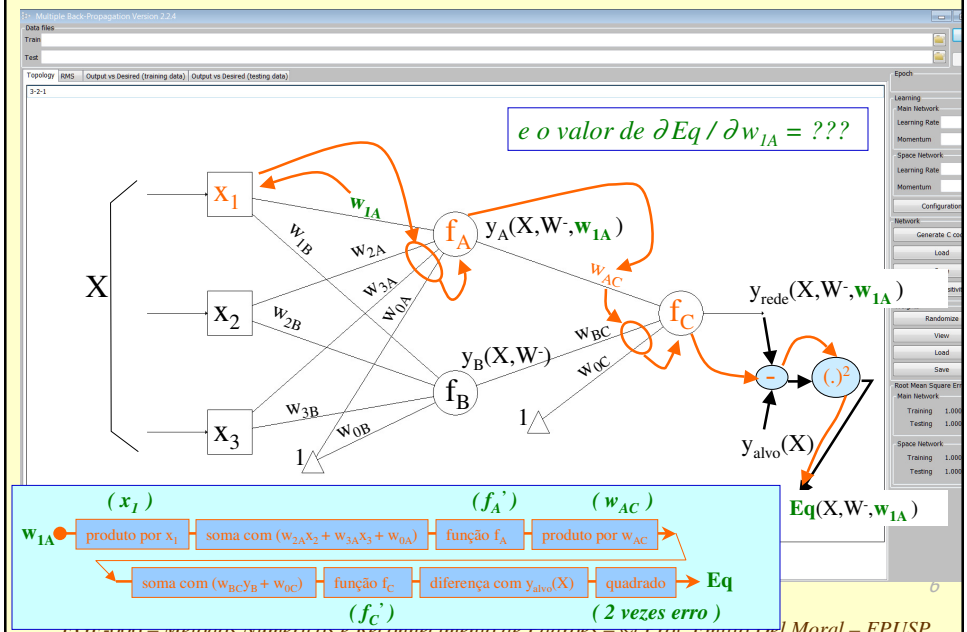
$$\mathbf{Grad(Eqm) =}$$

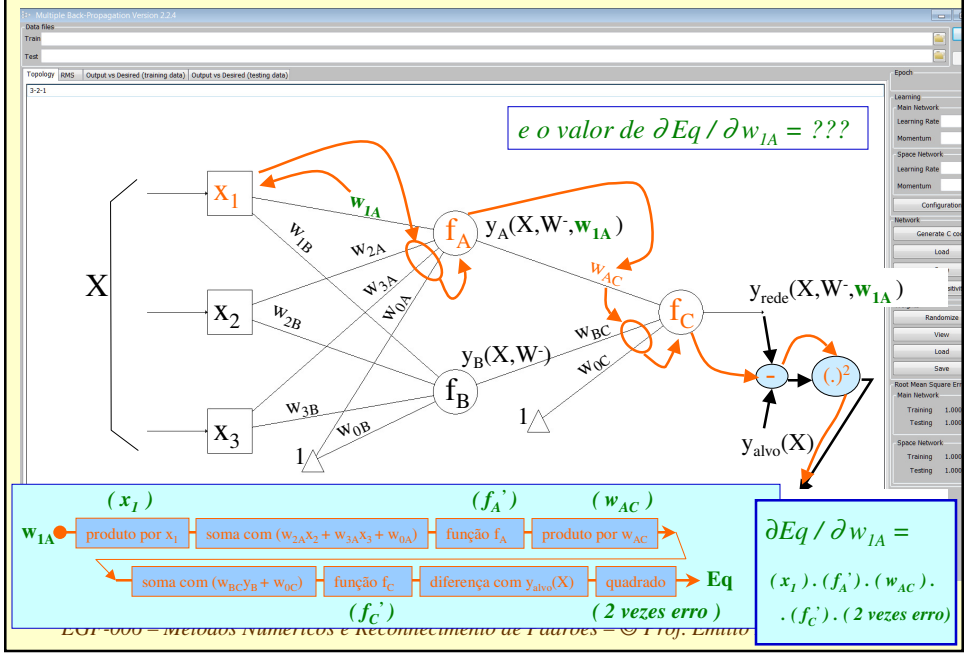
$$\mathbf{Grad(\Sigma_\mu\ Eq^\mu)\ /\ M}$$

$$\mathbf{\Sigma_\mu\ Grad\ (Eq^\mu)\ /\ M}$$




a cadeia de cálculos desde w_{1A}
... até ... Eq
X
x_1
x_2
x_3
w_{1A}
w_{1B}
w_{2A}
w_{3A}
w_{0A}
w_{2B}
w_{3B}
w_{0B}
f_A
f_B
f_C
$y_A(X,W^-,w_{1A})$
$y_B(X,W^-)$
w_{AC}
w_{BC}
w_{0C}
1
$y_{rede}(X,W^-,w_{1A})$
$y_{alvo}(X)$
$(.)^2$
$Eq(X,W^-,w_{1A})$
w_{1A}
produto por x_1
soma com $(w_{2A}x_2 + w_{3A}x_3 + w_{0A})$
função f_A
produto por w_{AC}
soma com $(w_{BC}y_B + w_{0C})$
função f_C
diferença com $y_{alvo}(X)$
quadrado
Eq


e o valor de $\partial Eq / \partial w_{1A}$ = ???
X
x_1
x_2
x_3
w_{1A}
w_{1B}
w_{2A}
w_{3A}
w_{0A}
w_{2B}
w_{3B}
w_{0B}
f_A
f_B
f_C
$y_A(X,W^-,w_{1A})$
$y_B(X,W^-)$
w_{AC}
w_{BC}
w_{0C}
1
$y_{rede}(X,W^-,w_{1A})$
$y_{alvo}(X)$
$(.)^2$
$Eq(X,W^-,w_{1A})$
(x_1)
(f_A')
(w_{AC})
w_{1A}
produto por x_1
soma com $(w_{2A}x_2 + w_{3A}x_3 + w_{0A})$
função f_A
produto por w_{AC}
soma com $(w_{BC}y_B + w_{0C})$
função f_C
diferença com $y_{alvo}(X)$
quadrado
Eq
(f_C')
(2 vezes erro)



Lembretes ....

- Na maioria dos slides anteriores, onde aparece X, leia-se X^{μ}, não incluído para não complicar demais os desenhos
- ... similarmente, onde aparece y_{alvo}, leia-se $y_{alvo}{}^{\mu}$. Idem para os Eq, leia-se Eq^{μ}
- Nos itens de cadeia de derivadas (f_A') e (f_C') , atenção para os valores dos argumentos, que devem ser os mesmos de f_A e f_C na cadeia original que leva w_{1A} a Eq.
- ... lembrando ... na cadeia original tinhamos ...
  - para f_C: $f_C(w_{AC} \cdot f_A(w_{1A}.x_1 + w_{2A}.x_2 + w_{0A}) + w_{BC} \cdot f_B(w_{1B}.x_1 + w_{2B}.x_2 + w_{0B}) + w_{0C})$
  - para f_A: $f_A(w_{1A}.x_1 + w_{2A}.x_2 + w_{0A})$
- Similarmente, para o bloco "quadrado", cuja derivada é a função "2 vezes", o argumento é $[y_{rede}(X,W) - y_{alvo}(X)]$

Lembretes ....

- O mesmo que foi feito para w_{1A} deve ser feito agora para os demais 10 pesos: w_{2A}, w_{3A}, w_{0A}, w_{1B}, w_{2B}, w_{3B}, w_{0B}, w_{AC}, w_{BC}, e w_{0C} !
- Assim compomos um um gradiente de 11 dimensões, com as derivadas de Eq^{μ} com relação aos 11 diferentes pesos w: $Grad_W (Eq^{\mu})$
- Essas 11 fórmulas devem ser aplicadas repetidamente aos M exemplares numéricos de X^{μ} e y_{alvo}^{μ}, calculando M gradientes!
- Com eles, se obtém o gradiente médio dos M pares empíricos: $Grad_W (Eqm) = [\Sigma_{\mu} Grad_W (Eq^{\mu})] / M$
- Esse gradiente <u>médio</u> é a Bussola do Gradiente!