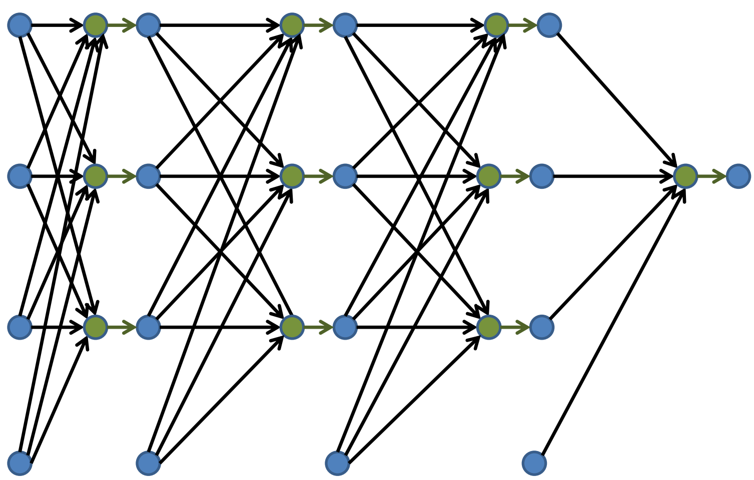
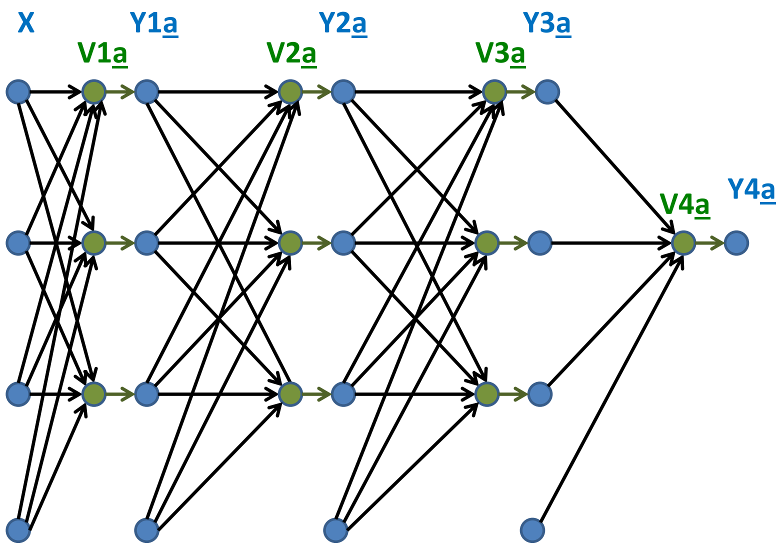
Processo de aproximações sucessivas ao Eqm mínimo:

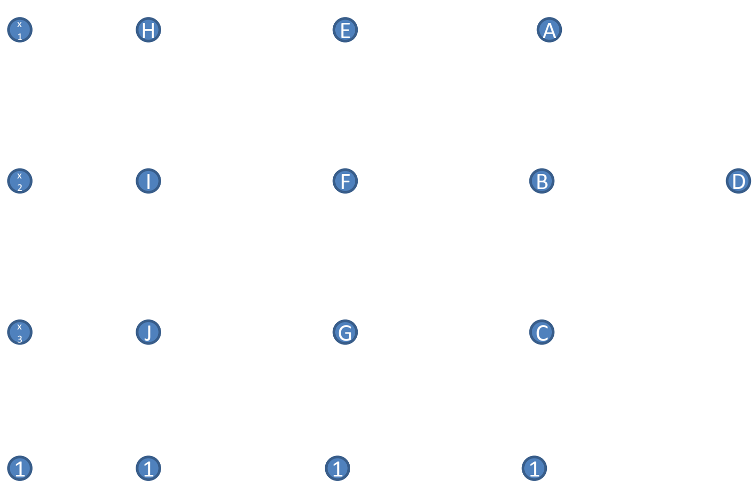
- Chute um vetor W inicial, e chame-o de “Wcorrente”, ou “W melhor até agora”
- Em loop EXTERNO, repita 1, 2 e 3 a seguir, até obter Eqm zero, ou Eqm baixo o suficiente, ou Eqm estável, ou estourar um número de adaptações ΔW limite:
  - 1) Determine o vetor gradiente do Eqm, nesse espaço de Ws. Essa determinação é feita através de um loop varrendo todos os M exemplos (X^{μ};y^{μ}). Cada passo desse loop INTERNO envolve 1.1, 1.2 e 1.3 como segue:
    - 1.1) Calcule o gradiente de Eq^{μ} associado a apenas um exemplo μ:
    - 1.2) Cada cálculo desses, associado a um μ apenas, envolve calcular tantas derivadas parciais de Eq^{μ} quantos pesos existam na rede. Isso exige primeiro calcular o valor do argumento de cada tangente hiperbólica e depois usar esses valores dos argumentos nas derivadas da regra da cadeia, necessárias ao cálculo das derivadas parciais de Eq^{μ} com relação aos vários pesos da rede;
    - 1.3) Vá varrendo μ (lembre que μ vai de 1 até M), e somando os gradientes obtidos para cada Eq^{μ}, para ir compondo o vetor gradiente de Eqm, que na verdade é a soma de todos os gradientes coletados para os diversos μ; saia deste loop INTERNO somente quando passar por todos os μ.
  - 2) Ao sair do Loop INTERNO acima, estamos prontos para dar um pequeno passo vetorial ΔW no espaço de pesos, com a direção e magnitude dados por

    $-\eta$. vetor gradiente de Eqm. Com isso, atualizamos / melhoramos o vetor Wcorrente
  - 3) Em seguida a dar tal passo ΔW, avalie se Eqm é pequeno / estável / decrescente, etc, e se não decidir parar o processo, prepare-se para um novo pequeno passo ΔW (volte ao passo do cálculo do gradiente – passo 1)

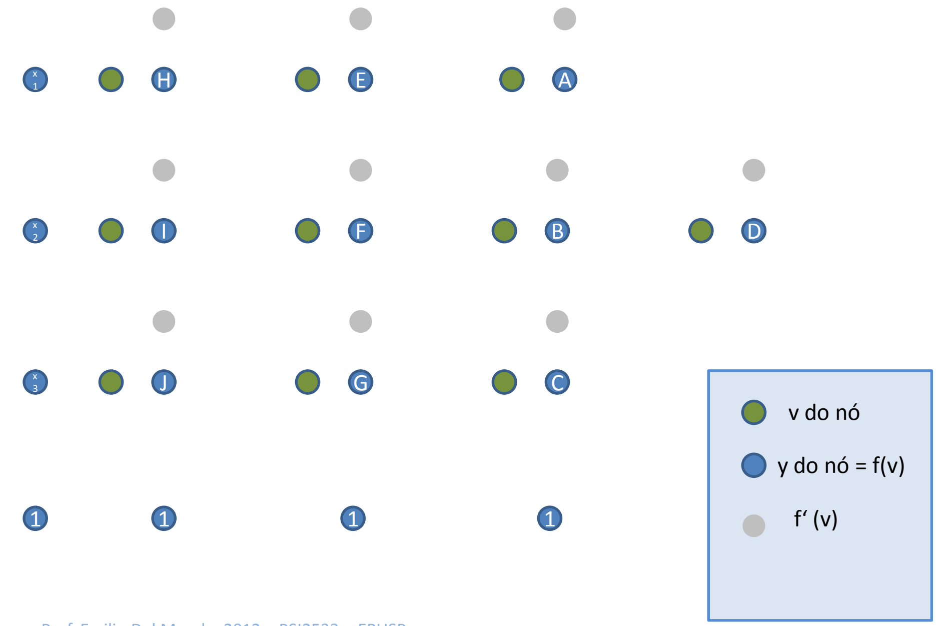




Destacando as variaveis x (entradas da rede) e y (saidas dos nós), e dando um nome a cada nó ...



Listando as variaveis relevantes no fluxo direto (... x, ... v, ... y) e incluindo f´ de cada no´.



Prof. Emilio Del Moral – 2012 – PSI2533 – EPUSP ·

... Relação entre gradiente descendente do erro e as derivadas parciais de y_{rede} com relação aos pesos sinápticos e com relação aos $y_{nó}$

$$\boxed{\Delta\vec{W} = -\eta.\vec{\nabla}Eqm = -\eta.\left(\frac{\partial Eqm}{\partial w_1}, \frac{\partial Eqm}{\partial w_2}, \frac{\partial Eqm}{\partial w_3} ...\right)}$$

$$\frac{\partial Eqm}{\partial w_{especifico}} =$$

$$= \frac{1}{M}\sum_{\mu=1}^{M}\frac{\partial Eq^{\mu}}{\partial w_{especifico}} = \frac{1}{M}\sum_{\mu=1}^{M}\left[\frac{\partial Eq^{\mu}}{\partial y_{rede}(\vec{X}^{\mu},W)}\right].\left[\frac{\partial y_{rede}(\vec{X}^{\mu},W)}{\partial w_{especifico}}\right]$$

$$= \frac{1}{M}\sum_{\mu=1}^{M}[2.(y_{rede}(\vec{X}^{\mu},W) - y^{\mu})].\left[\frac{\partial y_{rede}(\vec{X}^{\mu},W)}{\partial w_{especifico}}\right]$$

$$\boxed{= \frac{1}{M}\sum_{\mu=1}^{M}[2.(y_{rede}(\vec{X}^{\mu},W) - y^{\mu})].\left[\frac{\partial y_{rede}(\vec{X}^{\mu},W)}{\partial y_{nó}}\right].\left[\frac{\partial y_{nó}(\vec{X}^{\mu},W)}{\partial w_{especifico}}\right]}$$

Prof. Emilio Del Moral – 2012 – PSI2533 – EPUSP ·

Definição, e uma das interpretações do Retropropagador (∂ y_rede / ∂ y_nó)

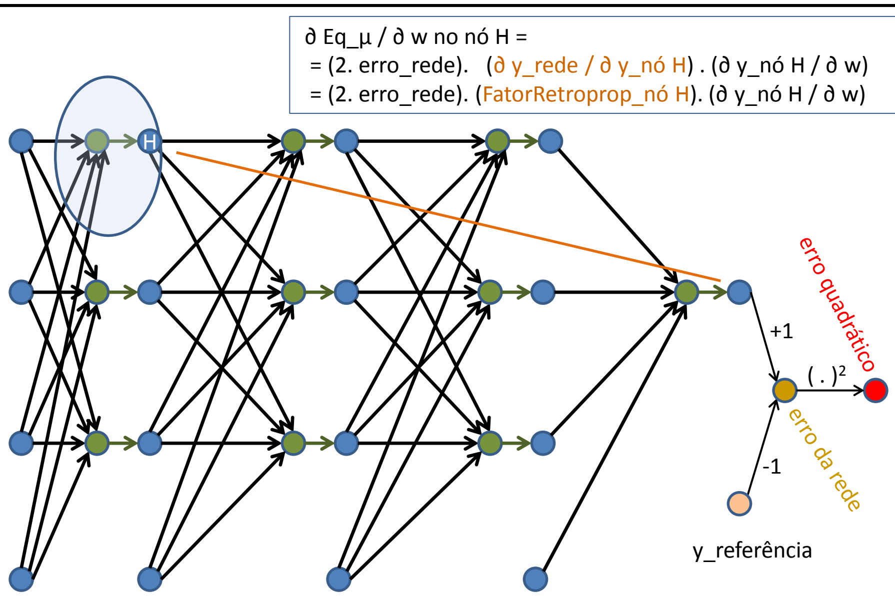
- Definição: (∂ y_rede / ∂ y_nó)
- Uma interpretação: ele nada mais é que o segundo "1/3" de uma cadeia de 3 passos, no cálculo de (∂ Eq_μ / ∂ w):

∂ Eq_μ / ∂ w =

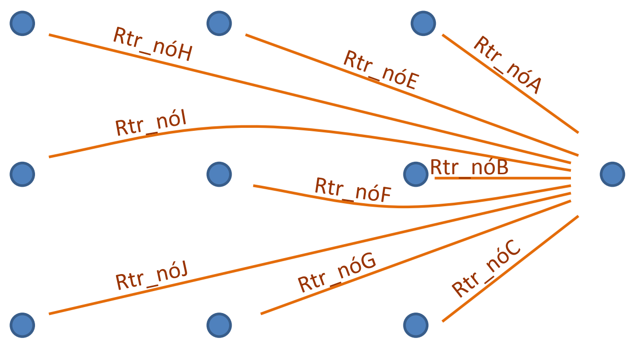
= (∂ Eq_μ / ∂ y_rede). (∂ y_rede / ∂ y_nó). (∂ y_nó / ∂ w)

= (2. erro_rede). (∂ y_rede / ∂ y_nó). (∂ y_nó / ∂ w)

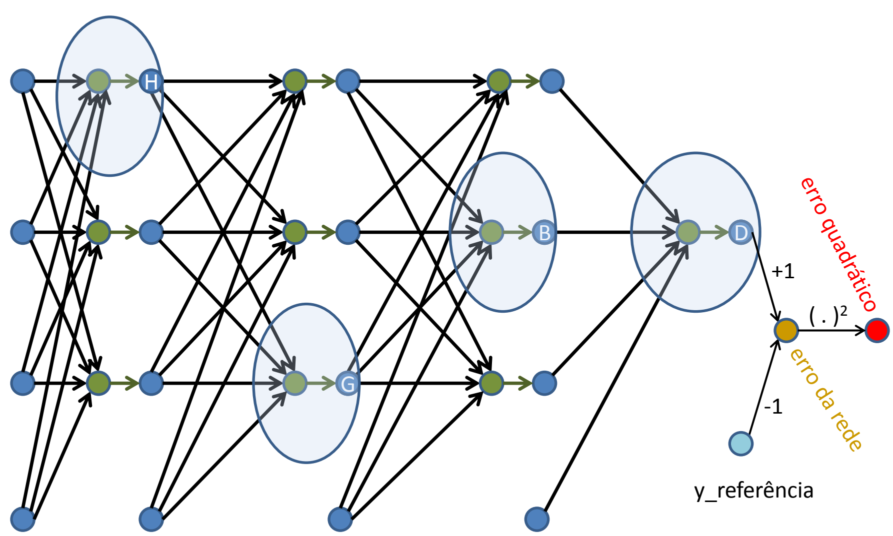
= (2. erro_rede). (FatorRetroprop_nó). (∂ y_nó / ∂ w)



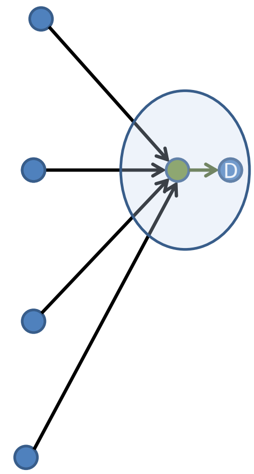
Temos um retropropagador para cada no´ da rede, a excessao do no´ de saida



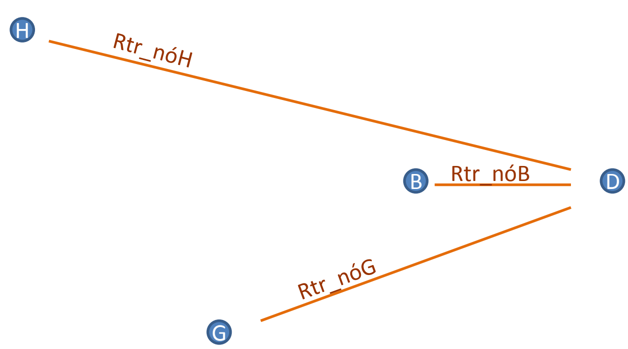
Escolhamos alguns nós significativos para ilustrar os retropropagadores: H, G, B, e D

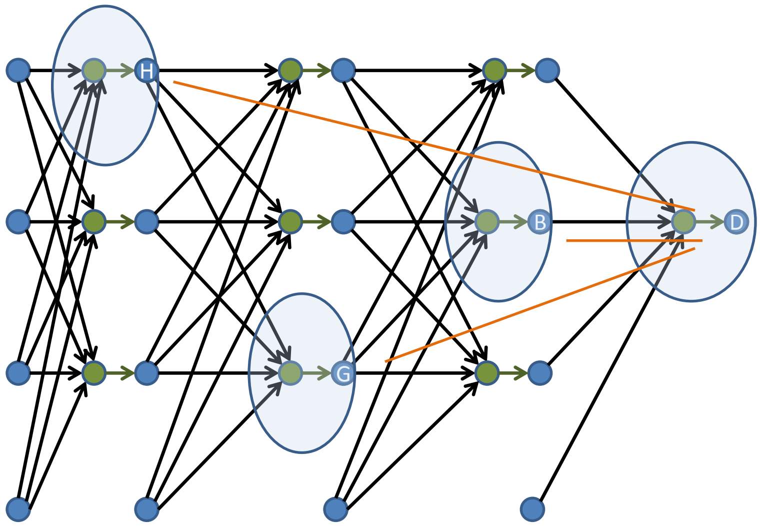


Para o no´ D, por ser o no´ de saida, não temos necessidade de retropropagador (dito de outra forma, y_nó = y_rede, ou ainda ... ∂ y_rede / ∂ y_nó =1)



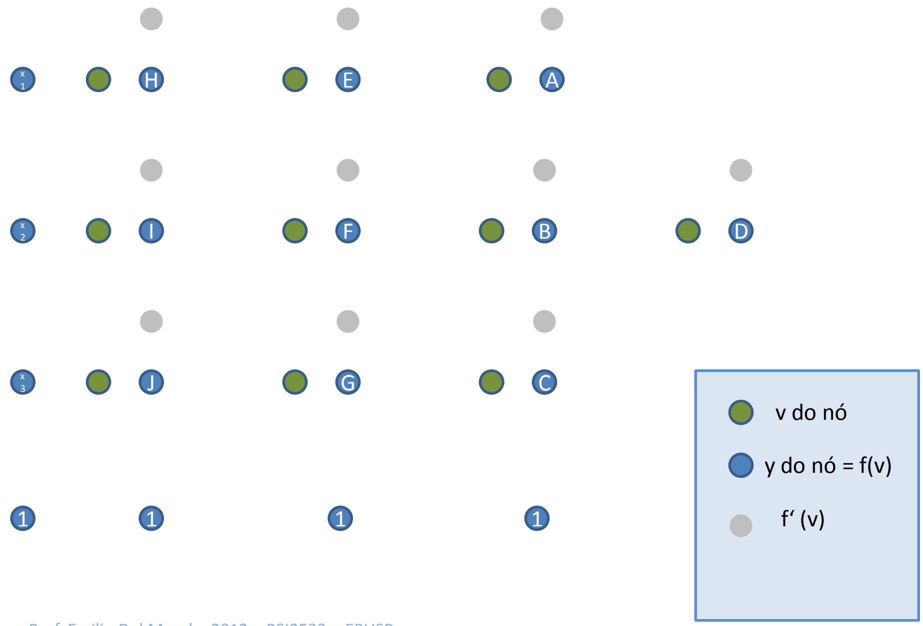
Já´ para os nós H, G, e B, que não são de saida (so´ o nó D e´ de saida), temos sim que usar o retropropagador



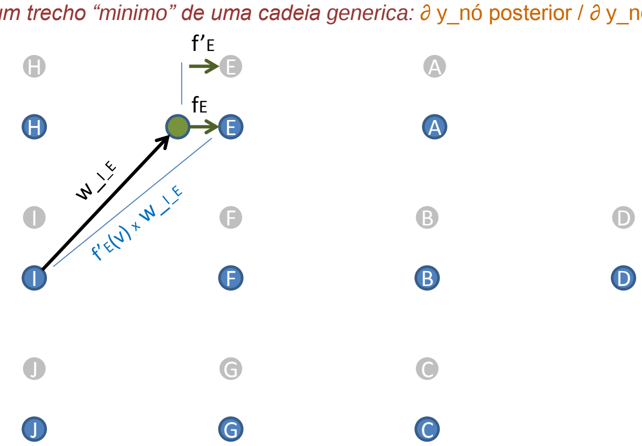


Calculando os retro-propagadores

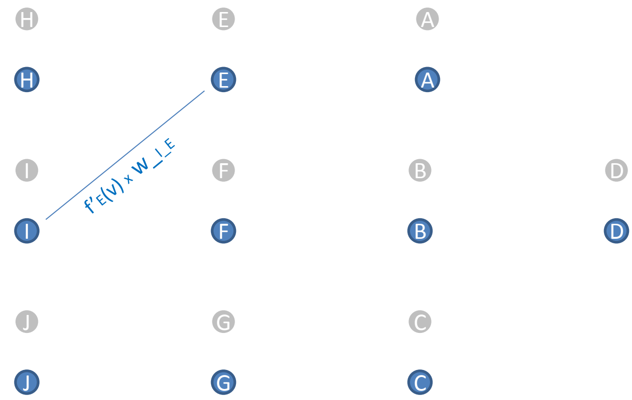
- Primeiro vamos preparar o terreno ...
- Derivadas parciais de y_rede com relação a y_nó são calculadas tendo como cálculos intermediários os elementos que compõem a cadeia que leva y_nó até y_rede
- Definamos mais claramente a natureza desses elementos intermediários da cadeia



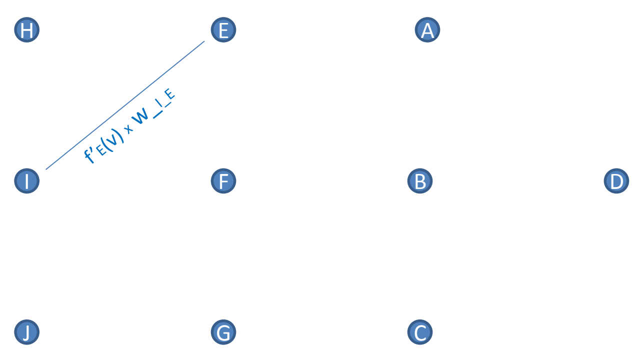
Calculemos um trecho “mimimo” de uma cadeia generica: ∂ y_nó posterior / ∂ y_nó anterior



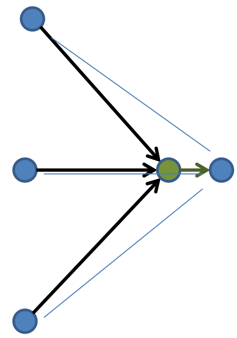
Façamos isso p/ o caso especifico dos nos I e E: ∂ y_nó posterior / ∂ y_nó anterior = w_I_E . f´(v)



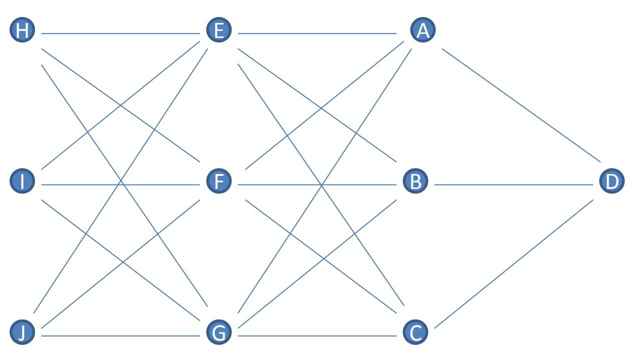
Façamos isso p/ o caso especifico dos nos I e E: ∂ y_nó posterior / ∂ y_nó anterior = w_I_E . f´(v)



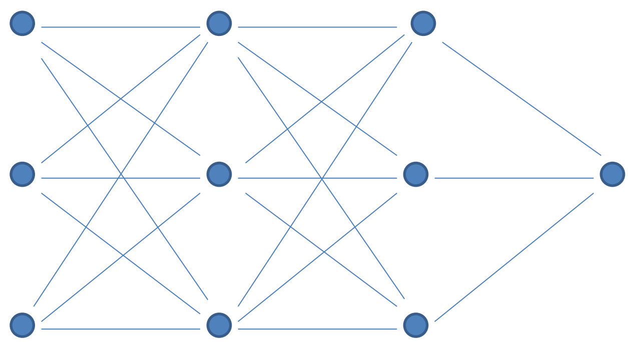
Podemos usar o conceito (∂ y_nó posterior / ∂ y_nó anterior = w . f´(v)) para outros nós na rede:



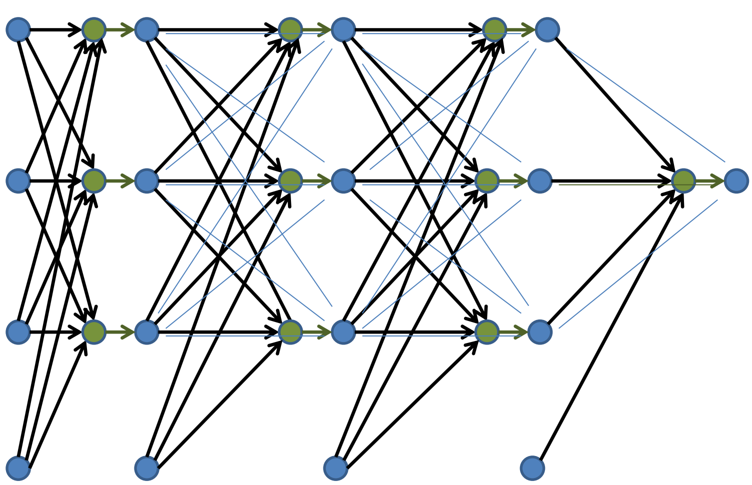
... e fazer o mesmo para todos os pares de nós anteriores / posteriores pertinentes:



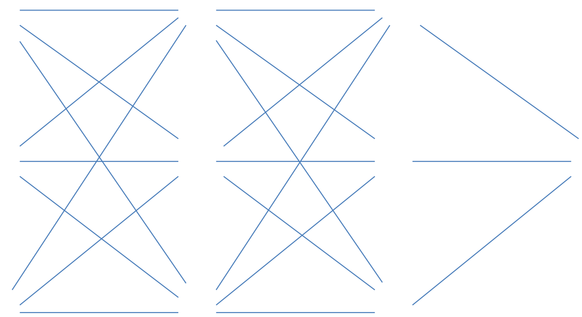
- Acima, cada círculo azul representa uma variável do tipo "y_de_nó";
- Cada linha azul representa uma "derivada de trecho", trecho esse que inclui tanto um peso w quanto a função f do nó que segue esse peso w. A linha azul que liga y_A com y_D, por exemplo, representa uma derivada de trecho com valor dado pelo produto (w_AD x f´D), e esse resultado de produto também corresponde ao valor da derivada parcial
(del y_D / del y_A)

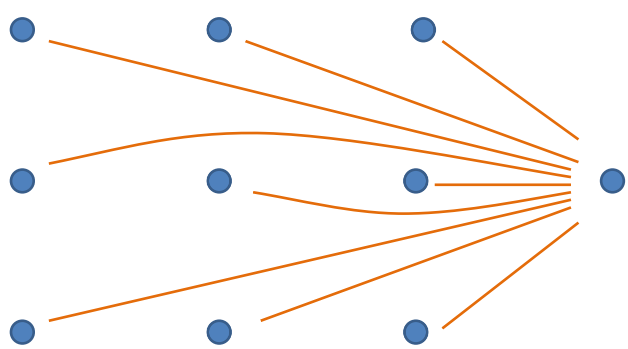


Grafo original da rede e trechos “ w.f’(v) ” em azul, sobrepostos ao grafo original:

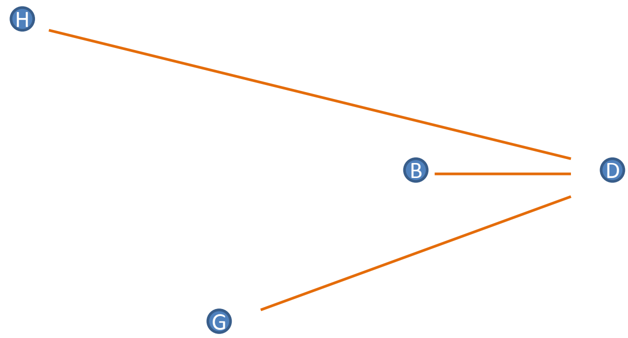


Agora, apenas os trechos " w.f'(v) " ... Eles podem ser calculados antecipadamente e depois usados inumeras vezes na composiçao de retropropagadores de nós genericos



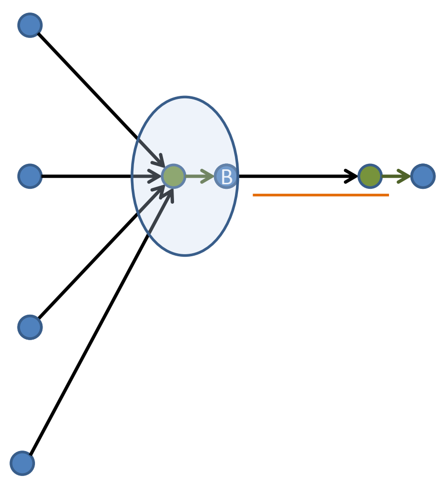


Voltemos no entanto aos 3 retropropagadores tomados como exemplo.
Vejamos como eles são relacionaveis com os trechos w.f'(v) pre´- calculados.
Ataquemos caso a caso ... Primeiro para o no´ B, depois G, depois H

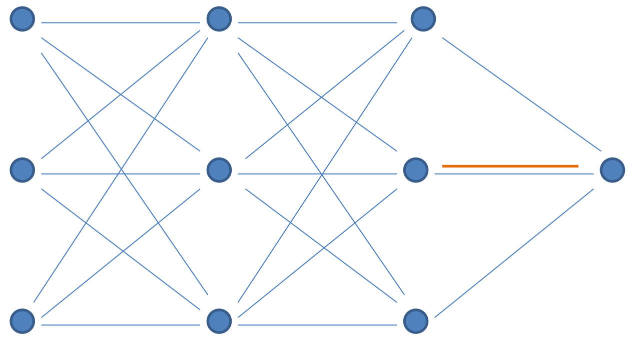


No caso do nó B:

∂ Eq_μ / ∂ w no nó B =
= (2. erro_rede). (∂ y_rede / ∂ y_nó B) . (∂ y_nó B / ∂ w)
= (2. erro_rede). (FatorRetroprop_nó B). (∂ y_nó B / ∂ w)



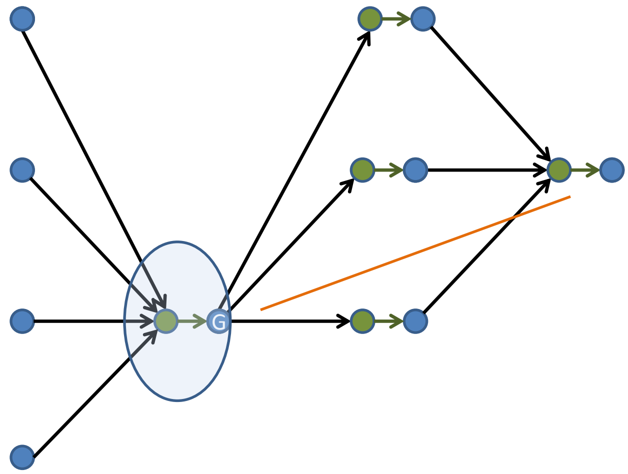
... o retropropagador coincide com um dos trechos w.f'(v)

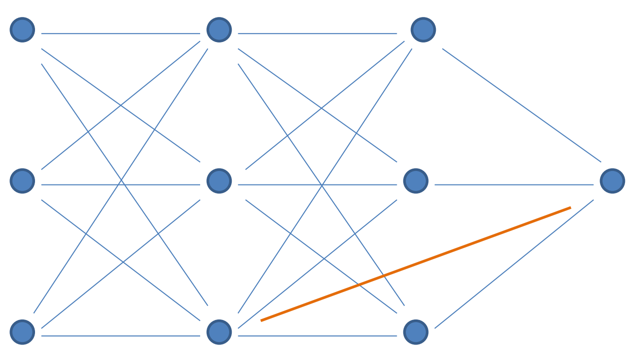


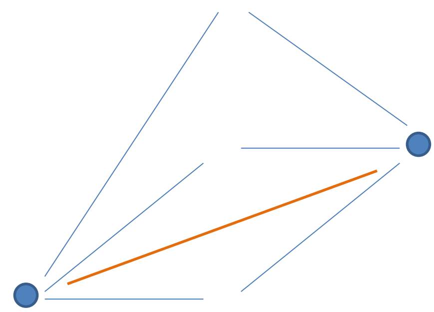
... o retropropagador coincide com um dos trechos w.f'(v)

No caso do nó G:

∂ Eq_µ / ∂ w no nó G =
= (2. erro_rede). (∂ y_rede / ∂ y_nó G) . (∂ y_nó G / ∂ w)
= (2. erro_rede). (FatorRetroprop_nó G). (∂ y_nó G / ∂ w)

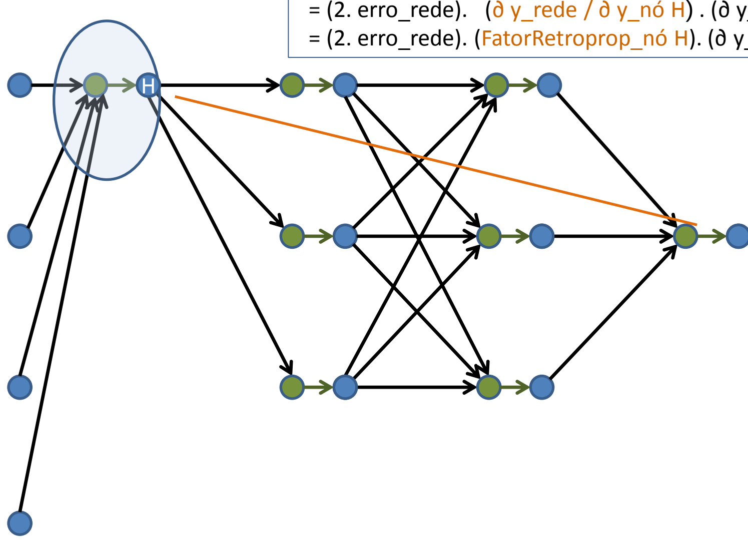


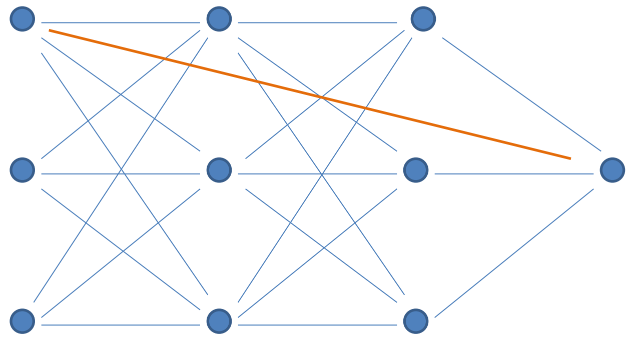


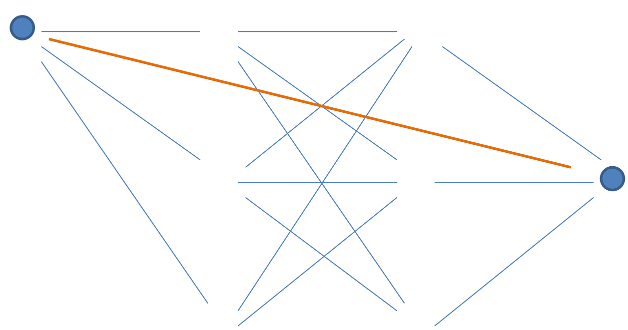


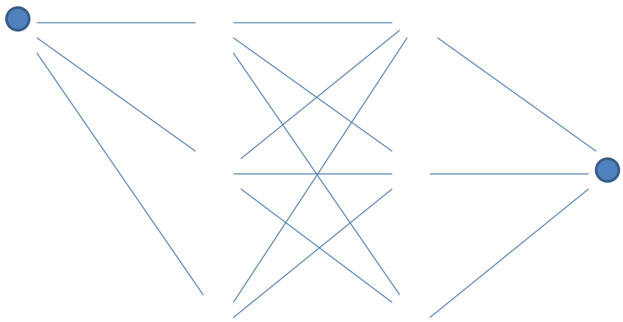
No caso do nó H:

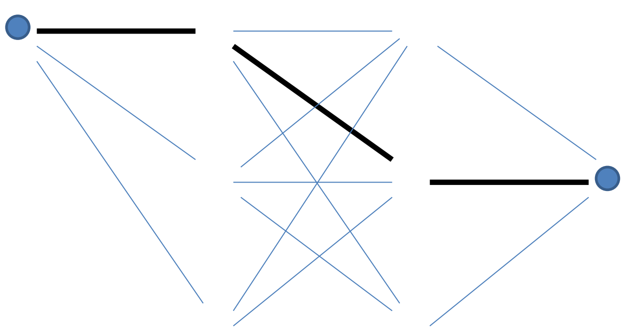
∂ Eq_µ / ∂ w no nó H =
= (2. erro_rede). (∂ y_rede / ∂ y_nó H) . (∂ y_nó H / ∂ w)
= (2. erro_rede). (FatorRetroprop_nó H). (∂ y_nó H / ∂ w)

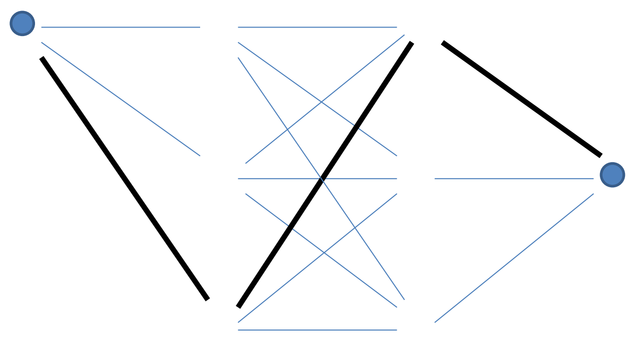


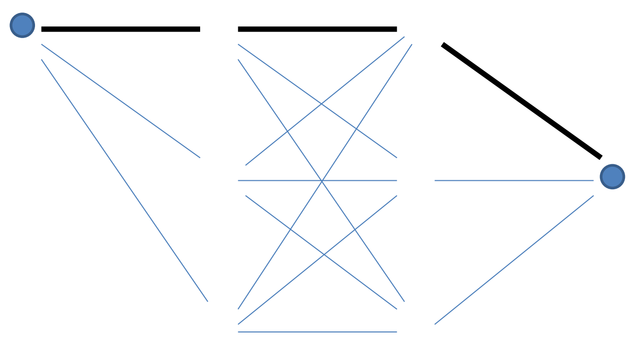


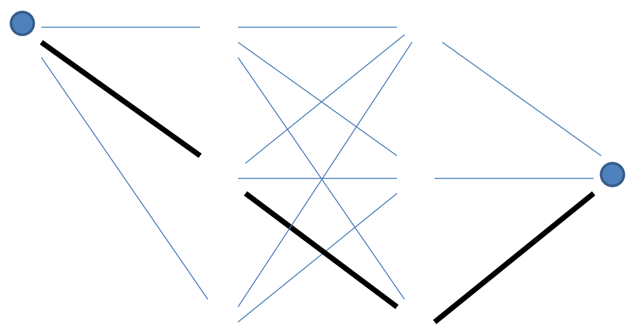


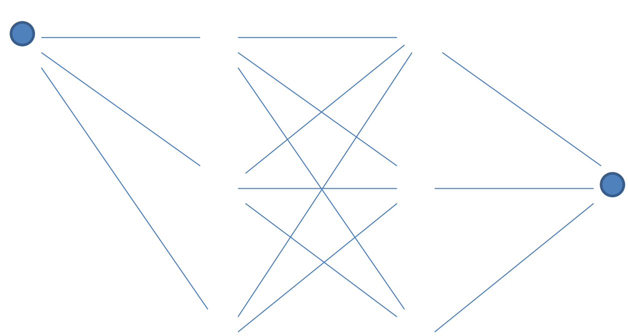






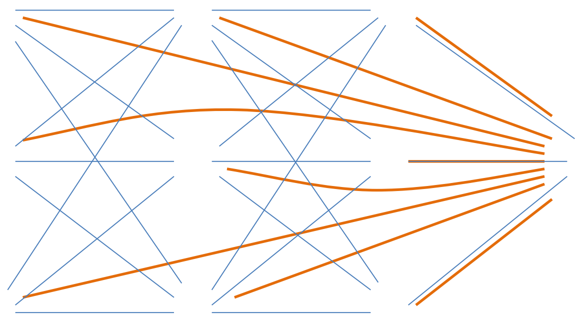


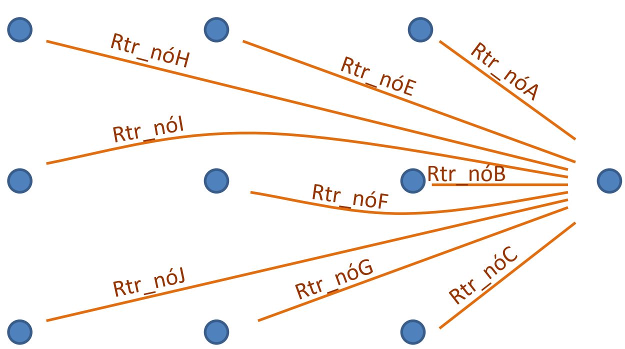




Total de 9 = 3*3*1 possíveis "caminhos que se bifurcam" (caminhos paralelos), que devem ser somados para compor a regra da cadeia

Em suma, com base nos 3 exemplos de nós abordados (B, G e H), fica claro que os trechos " w.f'(v) " em azul calculados antecipadamente são re-usados varias vezes na composiçao dos diversos retropropagadores de nós genericos





Na aula 8 …
Error Back Propagation, com a conceito de “ erro do no’ ”

Alternativa à retro-propagação de 1 passo: retro-propagação camada a camada

- Até o momento vimos como implementar o gradiente descendente com o conceito de retro-propagadores, que são calculados para cada nó individualmente e usados na seguinte expressão:

  ∂ Eq_μ / ∂ w = (∂ Eq_μ / ∂ y_rede). (∂ y_rede / ∂ y_nó). (∂ y_nó / ∂ w)
  = (2. erro_rede). (∂ y_rede / ∂ y_nó). (∂ y_nó / ∂ w)
  = (2. erro_rede). (FatorRetroprop_nó). (∂ y_nó / ∂ w)

- Veremos agora uma forma alternativa, com cálculos recursivos e apoio no conceito de erro retro-propagado em cada nó, com base na seguinte revisita a ∂ Eq_μ / ∂ w ...

  = (2. erro_rede). (FatorRetroprop_nó). (∂ y_nó / ∂ w)

  = (2. [erro_rede . FatorRetroprop_nó] . (∂ y_nó / ∂ w))

  Definição de erro de nó