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)

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 $y_{nó}$

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

$$\partial Eqm/\partial w_{específico} =$$

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

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

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

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)

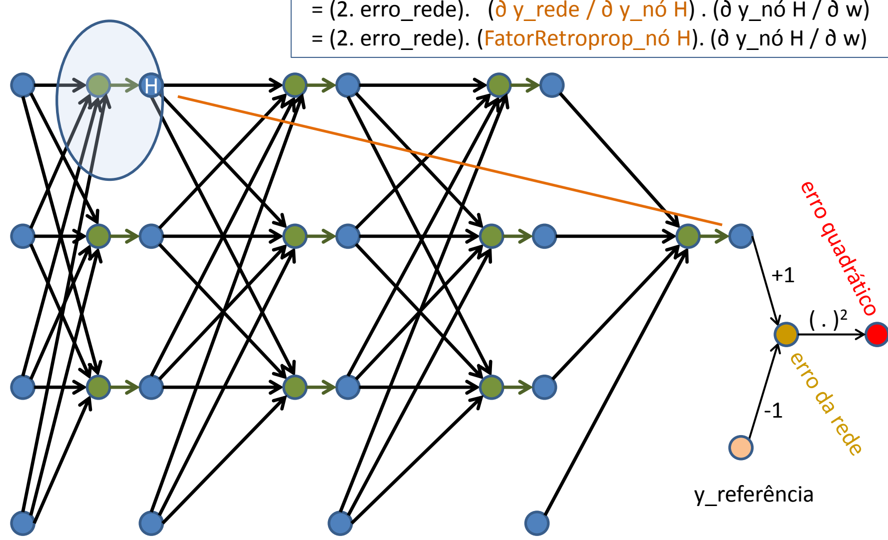
= (2. erro_rede). (∂ y_rede / ∂ y_nó). (∂ y_nó / ∂ w)

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

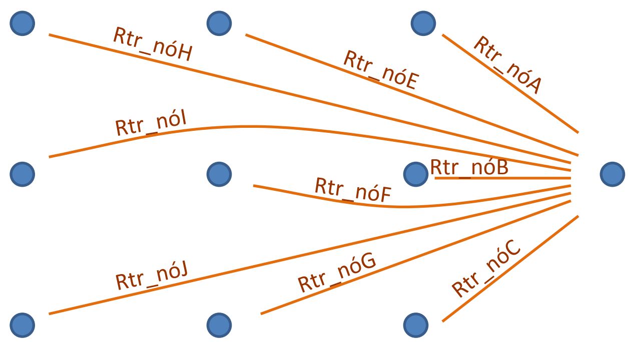
∂ Eq_μ / ∂ w no nó H =

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

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



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



ERROR BACK PROPAGATION

Algumas interpretações alternativas para o fator de retro-propagação de cada nó

- 1) 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)

- 2) Ele nada mais é que a derivada da saida da rede com relação a cada nó. Indica portanto, quando em produto juntamente com o fator (2 . erro_rede), como uma variação na saída do nó afeta o decréscimo do erro quadrático. Essa indicação do produto [FatorRetro_nó.2.(erro)] se dá tanto em direção (sinal) quanto em magnitude, já que é uma derivada (∂ Eq_μ / ∂ y_nó). O vetor de todos os retropopagadores portanto define a importância relativa (em magnitude e em sinal) de cada nó na formação do gradiente do erro e portanto a importância relativa entre nós na diminuição do Eq_μ, quando seus pesos são alterados

- 3) ... Esta fica clara adiante: Ele é um fator multiplicativo que permite transformar o erro da saída da rede num "erro local" de cada nó, sobre o qual a regra Delta local (regra do gradiente descendente para um nó) pode ser aplicada individualmente a cada nó, com erros diferenciados de maneira a manter adaptação concatenada na mesma direção do gradiente descendente conjunto. Isto ocorre porque ∂ Eq_μ / ∂ w ...
  = (2. erro_rede). (FatorRetroprop_nó). (∂ y_nó / ∂ w)
  = (2. [erro_rede. FatorRetroprop_nó] . (∂ y_nó / ∂ w))

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)

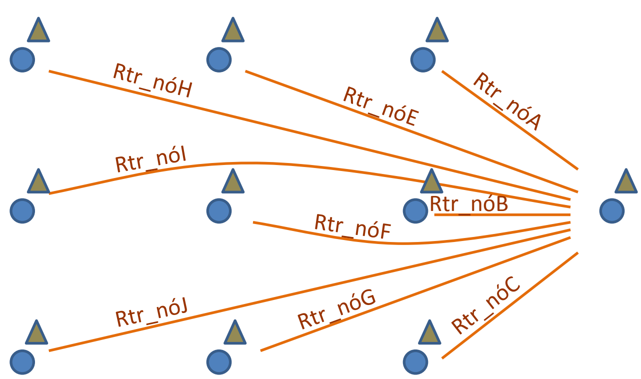
  = (2. [erro_rede. FatorRetroprop_nó] . (∂ y_nó / ∂ w))

  Definição de erro de nó

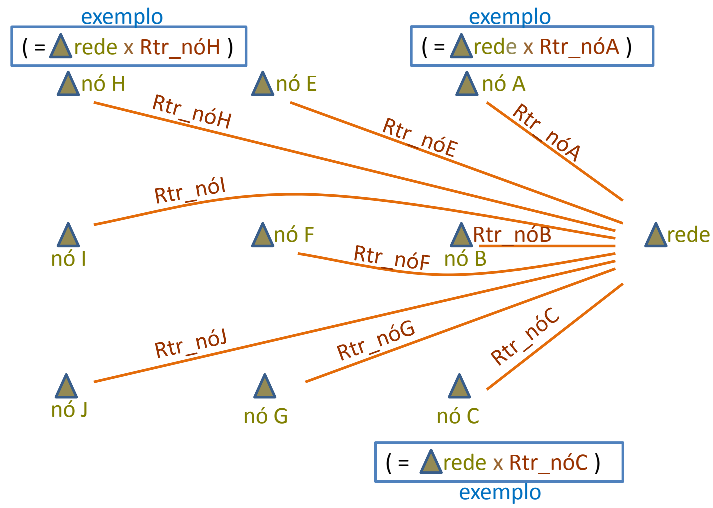
Definindo novas variáveis: uma nova variável auxiliar associada a cada nó, chamada erro do nó

erro do nó = [(erro da rede) x (retropagador do nó)]

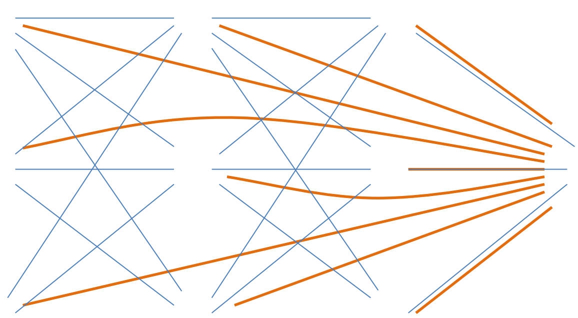
= [(y_rede-y_mu) x (del yrede / del y nó)]

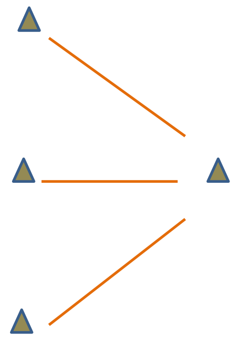


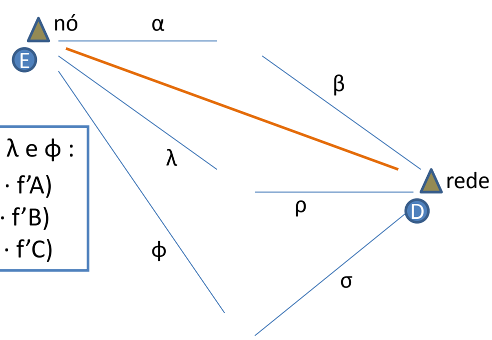
Os triângulos representam os erros associados a cada nó



De aula anterior, ficou claro que os trechos “ w.f’(v) ” em azul calculados antecipadamente são re-usados varias vezes na composiçao dos diversos retropropagadores de nós genericos







Lembrete da origem de α, λ e ϕ :

α : $\partial y\_A/\partial y\_E = (w\_E\_A \cdot f'A)$

λ : $\partial y\_B/\partial y\_E = (w\_E\_B \cdot f'B)$

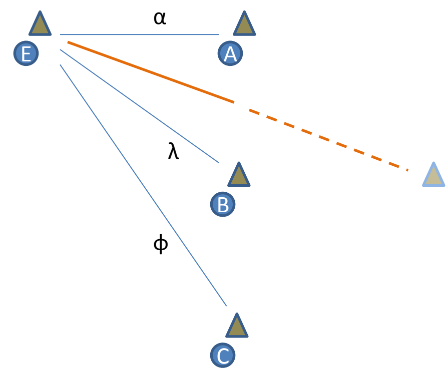
ϕ : $\partial y\_C/\partial y\_E = (w\_E\_C \cdot f'C)$

Lembrete da origem de β, ρ e σ:

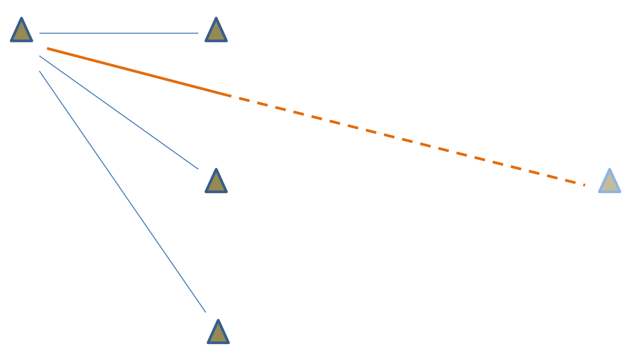
β : $\partial y\_D/\partial y\_A = (w\_A\_D \cdot f'D)$

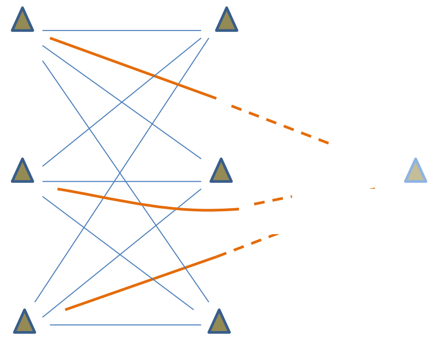
ρ : $\partial y\_D/\partial y\_B = (w\_B\_D \cdot f'D)$

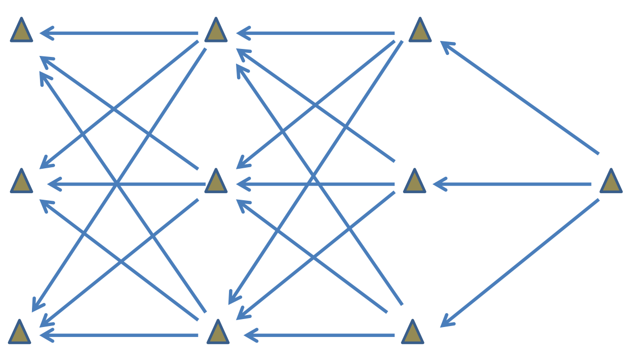
σ : $\partial y\_D/\partial y\_C = (w\_C\_D \cdot f'D)$



$$\text{erro E} = \alpha\cdot(\text{erro da rede} \cdot \beta) + \lambda\cdot(\text{erro da rede . P}) + \phi\cdot(\text{erro da rede . }\sigma)$$
$$= \alpha\cdot(\text{erro A}) + \lambda\cdot(\text{erro B}) + \phi\cdot(\text{erro C})$$





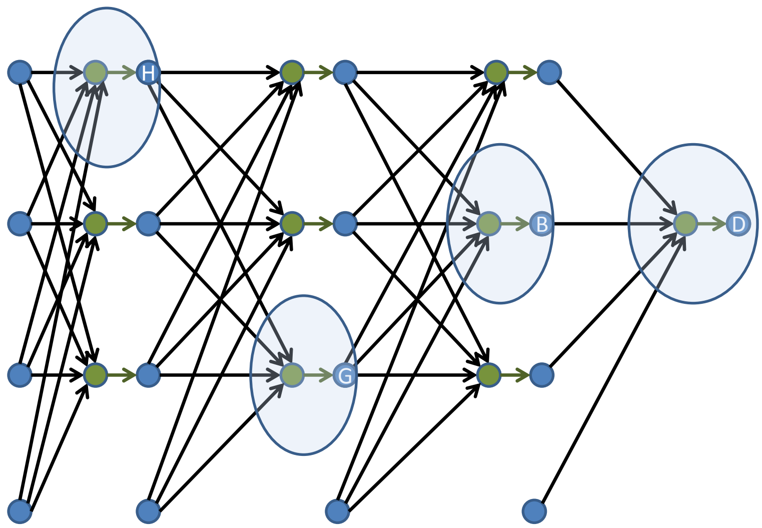


Resumo do EBP em 3 passos

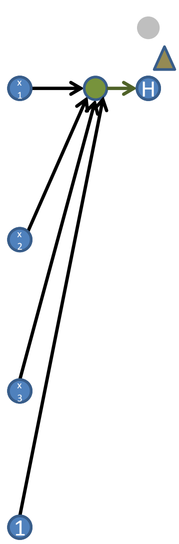
- Passo 1 - de Fluxo Direto:
  - Cálculo de V, Y, Y´, camada a camada, da 1ª à última
  - Cálculo de erro da rede, y(X_mu) - y_mu
  - Cálculos das "derivadas de trecho": [w_nó1_nó2 x f'nó2(v)]
- Passo 2 - de Fluxo Reverso:
  - Cálculo dos erros de nós ("deltas"), camada a camada, da última para a primeira, com o uso da técnica de retro-propagação em passos, camada a camada
- Passo 3 - de Cálculo de Adaptação de Pesos para o exemplar mu
  - Aplicar a regra Delta Clássica (de 1 nó), em paralelo = calculando simultaneamente as adaptações de todos os nós
  - Cada aplicação local da regra cássica envolve: saídas Y da camada anterior, erro no nó (delta retro-propagado) e f'(v) do nó
  - Guarde em memória o Delta W_mu "recomendado pelo exemplar um sendo analidado", sem adaptar pesos ainda!!! A adaptação só ocorrerá com a média total das recomendações individuais dadas por todos os M exemplares de treino.

Passo 3 (tambem reverso, complementando o passo 2)

Regra Delta Clássica, aplicada nó a nó, em toda a rede



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

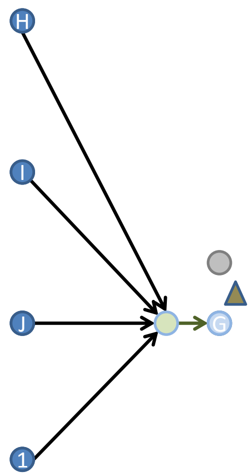


Elementos Importantes para o cálculo de Delta W do nó, segundo a regra Delta:

Entradas do nó, x_1, x_2, x_3, erro_nó, e f'(v) no nó

Os valores da saída y do nó e de v do nó não são necessários (a menos é claro para os cálculos previamente realizados, do erro e de f'(v) , na fase direta)

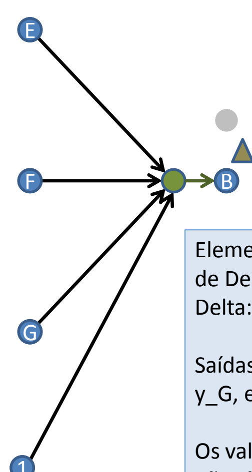
Prof. Emilio Del Moral – 2012 – PSI2533 – EPUSP ·



Elementos Importantes para o cálculo de Delta W do nó, segundo a regra Delta:

Saídas da camada anterior, y_H, y_I, y_J, erro_nó, e f'(v) no nó

Os valores da saída y do nó e de v do nó não são necessários (a menos é claro para os cálculos previamente realizados, do erro e de f'(v), na fase direta)



Elementos Importantes para o cálculo de Delta W do nó, segundo a regra Delta:

Saídas da camada anterior, y_E, y_F, y_G, erro_nó, e f'(v) no nó

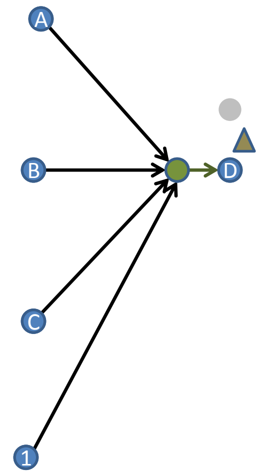
Os valores da saída y do nó e de v do nó não são necessários (a menos é claro para os cálculos previamente realizados, do erro e de f'(v), na fase direta)

Elementos Importantes para o cálculo de Delta W do nó, segundo a regra Delta:

Saídas da camada anterior, y_A, y_B, y_C, erro_nó, e f'(v) no nó.

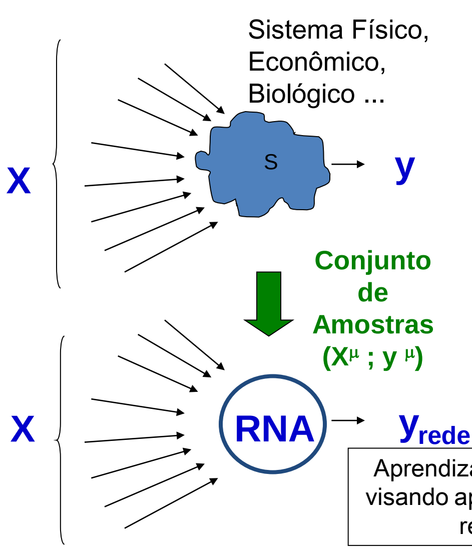
Neste nó específico, por ser o nó de saída, temos que o erro do nó coincide com o erro da rede (não se trata de erro retro-propagado)

Os valores da saída y do nó e de v do nó não são necessários (a menos é claro para os cálculos previamente realizados, do erro e de f'(v) , na fase direta)



Retornando `as nossas motivaçoes inincIais no calculo de adaptaçoes de pesos ...

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

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

$$\overrightarrow{\Delta W} = -\eta.\overrightarrow{\nabla Eqm}$$

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