

Mas cuidado ... você experimentou rodar mais de uma vez o MBP sobre os mesmos dados de treino e ver se o resultado final é o mesmo?

- Não há garantia de que o W randômico, seguido de Gradiente Descendente leve ao mínimo global quando há vários mínimos, só há garantia de que levará a um dos mínimos locais, que pode não ser o mínimo global
- Isto faz necessário rodar várias vezes a otimização de pesos e selecionar a configuração com melhor E_{qm}
- É legítimo ficarmos com a rede otimizada de menos E_{qm} final? Ou deveríamos ter um ataque de avaliar o E_{qm} médio de várias tentativas?
- *Outro assunto ... e quanto à variação associada ao k-fold Cross Validation? Média ou melhor E_{qm} ?*

134

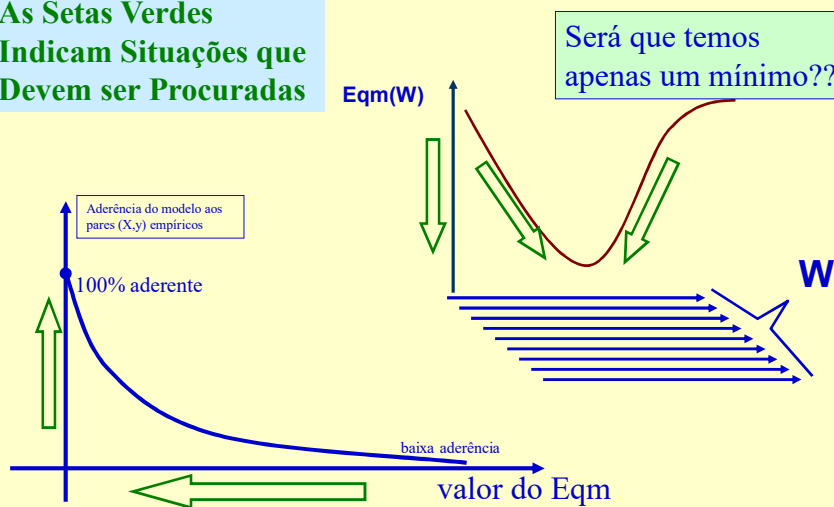
© Prof. Emilio Del Moral – EPUSP

134

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 E_{qm} possível

As Setas Verdes Indicam Situações que Devem ser Procuradas



135

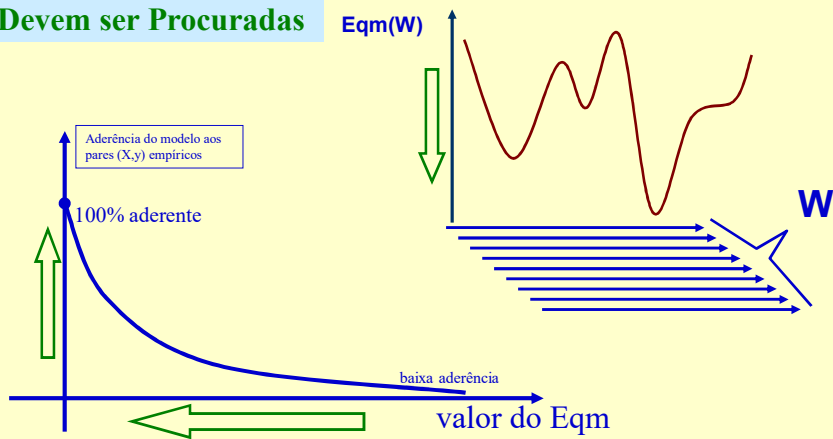
© Prof. Emilio Del Moral – EPUSP

135

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

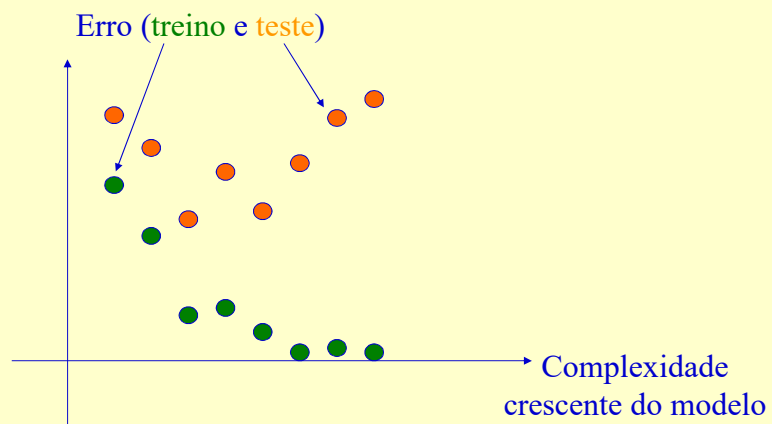
As Setas Verdes Indicam Situações que Devem ser Procuradas



© Prof. Emilio Del Moral – EPUSP

136

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) ...



© Prof. Emilio Del Moral – EPUSP

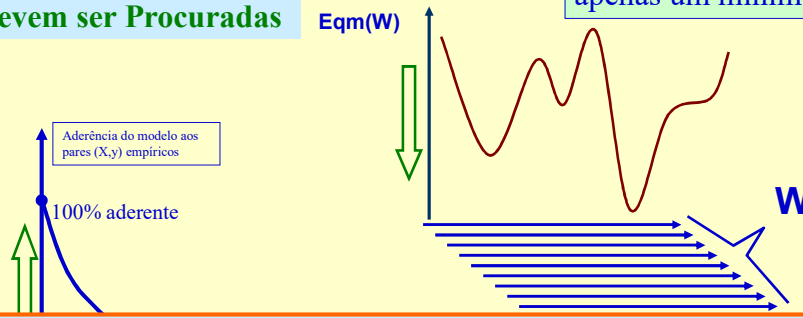
137

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 E_{qm} possível

As Setas Verdes Indicam Situações que Devem ser Procuradas

Será que temos apenas um mínimo??



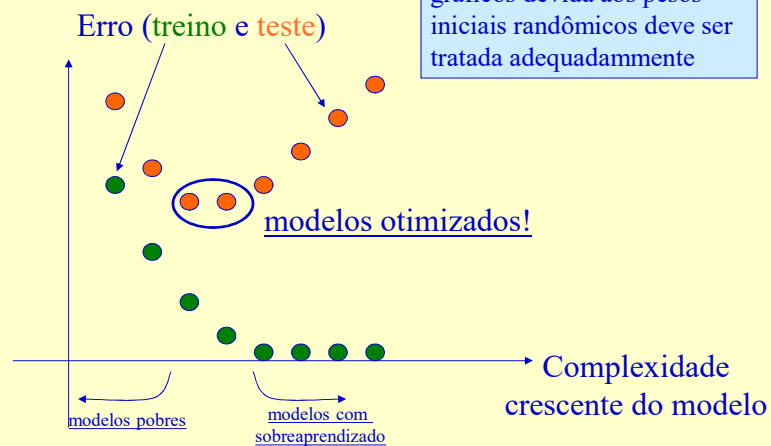
... Para não sermos reféns de mínimos locais com alto E_{qm} , podemos aplicar o gradiente descendente repetidamente na mesma RNA, com novos pesos iniciais randômicos em cada rodada, mantendo para o modelo final apenas os valores de pesos associados ao ensaio com o melhor dos resultados finais no E_{qm}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

© Prof. Emilio Del Moral – EPUSP

138

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



139

© Prof. Emilio Del Moral – EPUSP

139