Otimização da Estrutura - Poda

Algoritmos destrutivos – Prunning ou Poda

Inspeção

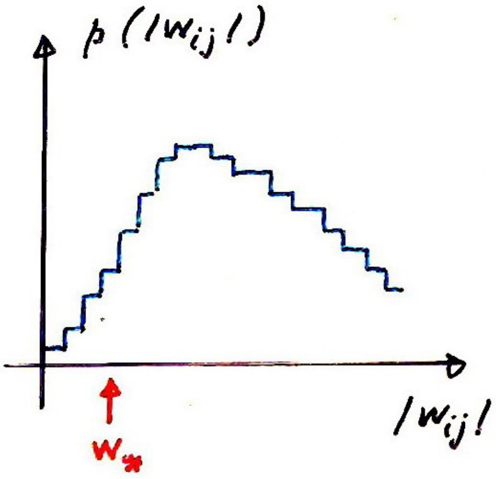
Treinar a rede (sem saturar – paralizar)

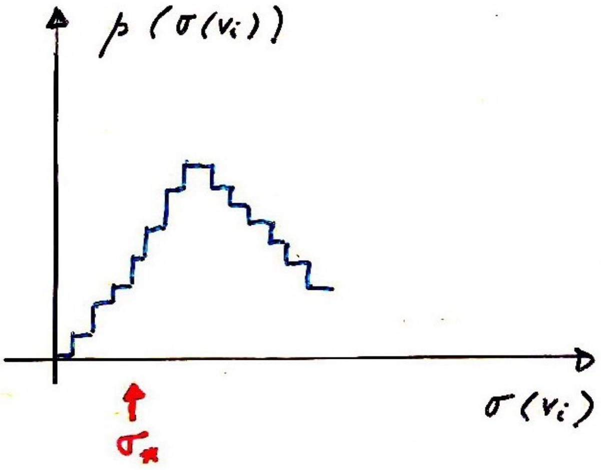
Eliminar

- **Sinapses pequenas** $|w| < w_{min}$
- **Neurônios da camada intermediárias com pouca variação na saída** $\sigma(v) < \sigma_{min}$

Retreinar para "ajuste fino"

Retomar a eliminação





Pouco eficaz,

a rede "distribui as tarefas" entre os neurônios

Alterando a função objetivo

$$F = F_0 + \gamma F_1 \qquad \gamma > 0$$

F_0 reduz o erro na saída **F_1 reduz a estrutura da rede**

Otimização de uma soma

$$\Delta \underline{w} \quad \Rightarrow \quad \Delta F = \Delta F_0 + \gamma\, \Delta F_1 < 0$$

$$\Rightarrow \begin{cases} \Delta F_0 < 0 & \Delta F_1 < 0 \\ \Delta F_0 < 0 & \Delta F_1 > 0 \\ \Delta F_0 > 0 & \Delta F_1 < 0 \end{cases}$$

Reduzindo sinapses

$$F = F_0 + \gamma F_1 = F_0 + \gamma \sum_{\forall i,j} w_{ij}^2$$

$$\frac{\partial F}{\partial w_{ij}} = \frac{\partial F_0}{\partial w_{ij}} + 2\gamma w_{ij}$$

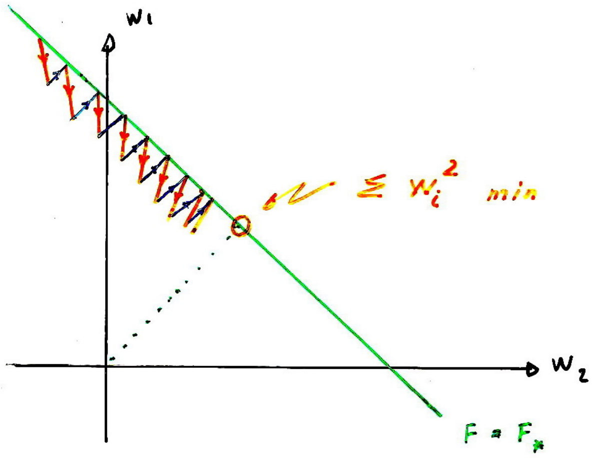
$$\Delta w_{ij} = -\alpha \frac{\partial F}{\partial w_{ij}} = -\alpha \frac{\partial F_0}{\partial w_{ij}} - 2\alpha\gamma w_{ij} = \Delta w_{ij\,BP} - 2\alpha\gamma w_{ij}$$

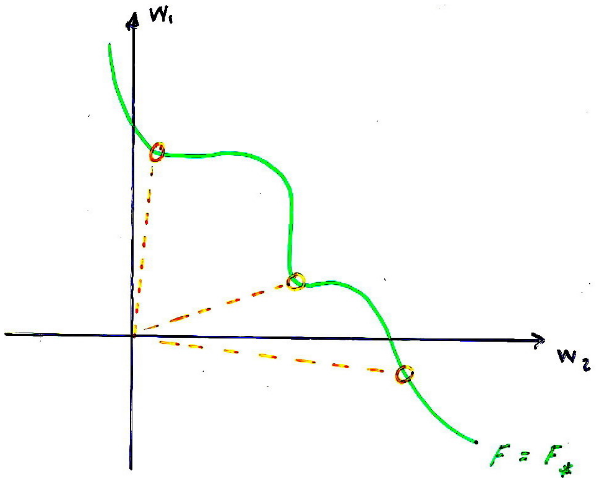
$$w_{ij\,novo} = (1 - 2\alpha\gamma) w_{ij\,antigo} + \Delta w_{ij\,BP} \qquad 0 < \gamma << \frac{1}{2\alpha}$$

Obs: sem redução de sinapses $w_{ij\,novo} = w_{ij\,antigo} + \Delta w_{ij\,BP}$

quando $|w_{ij}| < \min \;\Rightarrow$ eliminar $w_{ij} \;\Rightarrow\; w_{ij} = 0$

Evolução geométrica do processo





Critério para γ

$$F = F_0 + \gamma F_1 \qquad \begin{cases} \gamma \ll & \text{não reduz sinapses} \\ \gamma \gg & \text{não minimiza } F_0 \end{cases}$$

Heurística para γ

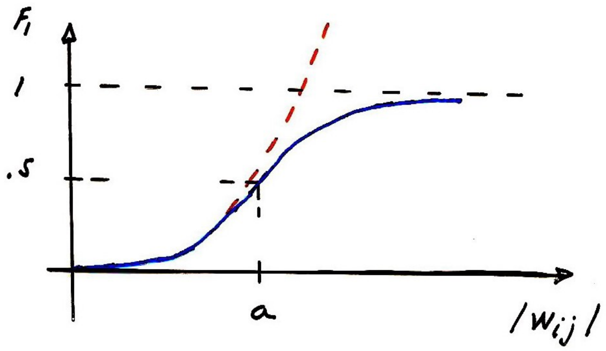
$$\left|\underline{\Delta w}_{F_0}\right| \approx \left|\underline{\Delta w}_{\gamma F_1}\right| = 2\alpha\gamma\left|\underline{w}\right|$$

$$\gamma \approx \begin{cases} \text{por valor médio} \quad \gamma \approx \dfrac{E\left(\left|\underline{\Delta w}_{F_0}\right|\right)}{2\alpha E\left(\left|\underline{w}\right|\right)} \\ \\ \text{por sinapse} \quad \gamma_{ij} \approx \dfrac{\left|\Delta w_{ij}\right|}{2\alpha\left|w_{ij}\right|} \end{cases}$$

Reduzindo as menores sinapses

$$F = F_0 + \gamma F_1 = F_0 + \gamma \sum_{\forall i,j} \frac{w_{ij}^2}{a^2 + w_{ij}^2} = F_0 + \gamma \sum_{\forall i,j} \frac{1}{a^2 + w_{ij}^2} w_{ij}^2$$

$$\left| w_{ij} \right| \begin{cases} \ll a & F_1 \simeq \dfrac{1}{a^2} w_{ij}^2 \\ \gg a & F_1 \simeq 1 \end{cases}$$



heurística: $a \approx \underset{\forall i,j}{E}\left(\left| w_{ij} \right| \right)$

$$F = F_0 + \gamma F_1 = F_0 + \gamma \sum_{\forall i,j} \frac{w_{ij}^2}{a^2 + w_{ij}^2} = F_0 + \gamma \sum_{\forall i,j} \frac{1}{a^2 + w_{ij}^2} w_{ij}^2$$

$$\frac{\partial F}{\partial w_{ij}} = \frac{\partial F_0}{\partial w_{ij}} + 2\gamma \frac{a^2 w_{ij}}{\left(a^2 + w_{ij}^2 \right)^2}$$

$$\Delta w_{ij} = -\alpha \frac{\partial F}{\partial w_{ij}} = -\alpha \frac{\partial F_0}{\partial w_{ij}} - 2\alpha\gamma \frac{a^2 w_{ij}}{\left(a^2 + w_{ij}^2 \right)^2} = \Delta w_{ij\,BP} - 2\alpha\gamma \frac{a^2}{\left(a^2 + w_{ij}^2 \right)^2} w_{ij}$$

$$w_{ij\,novo} = (1 - 2\alpha\gamma \frac{a^2}{\left(a^2 + w_{ij}^2 \right)^2}) w_{ij\,antigo} + \Delta w_{ij\,BP}$$

Obs: sem redução de sinapses $w_{ij\,novo} = w_{ij\,antigo} + \Delta w_{ij\,BP}$

quando $\left| w_{ij} \right| < \min$ $\Rightarrow$ eliminar w_{ij} $\Rightarrow$ $w_{ij} = 0$

Heurística para γ

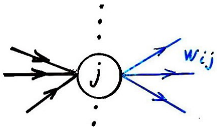
$$E\left|\Delta w_{ij\,BP}\right| \approx E\left|\Delta w_{ij\;\gamma F}\right|$$

$$E\left|\Delta w_{ij\,BP}\right| \approx 2\alpha\gamma \frac{a^2}{\left(a^2 + w_{ij}^2\right)^2} E\left|w_{ij}\right| = \frac{\alpha\gamma}{2a}$$

$$\gamma \approx \frac{2a}{\alpha} E\left|\Delta w_{ij\,BP}\right|$$

Eliminando neurônios da camada intermediária

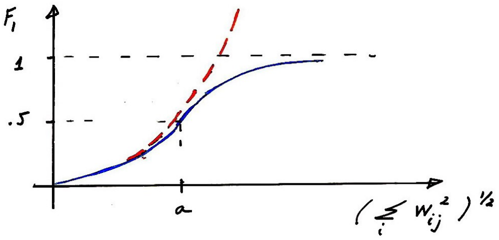
$$F = F_0 + \gamma F_1 = F_0 + \gamma \sum_{\forall j} \frac{\sum_{\forall i} w_{ij}^2}{a^2 + \sum_{\forall i} w_{ij}^2}$$



$\sum_{\forall j}$ somatório para todos os neurônios j da camada intermediária

$\sum_{\forall i} w_{ij}^2$ somatório para todas as sinapses do neurônio j da camada intermediária

heurística: $a \approx \left(\underset{\forall j}{E} \left(\sum_{\forall i} w_{ij}^2 \right) \right)^{1/2}$



$$F = F_0 + \gamma F_1 = F_0 + \gamma \sum_{\forall j} \frac{\sum_{\forall i} w_{ij}^2}{a^2 + \sum_{\forall i} w_{ij}^2} \qquad \frac{\partial F}{\partial w_{ij}} = \frac{\partial F_0}{\partial w_{ij}} + 2\gamma \frac{a^2 w_{ij}}{\left(a^2 + \sum_{\forall i} w_{ij}^2\right)^2}$$

$$\Delta w_{ij} = -\alpha \frac{\partial F}{\partial w_{ij}} = \Delta w_{ij\,BP} - 2\alpha\gamma \frac{a^2}{\left(a^2 + \sum_{\forall i} w_{ij}^2\right)^2} w_{ij}$$

$$w_{ij\,novo} = (1 - 2\alpha\gamma \frac{a^2}{\left(a^2 + \sum_{\forall i} w_{ij}^2\right)^2}) w_{ij\,antigo} + \Delta w_{ij\,BP}$$

Obs: sem redução de sinapses $w_{ij\,novo} = w_{ij\,antigo} + \Delta w_{ij\,BP}$

quando $\left| \sum_{\forall i} w_{ij}^2 \right| < \min \quad \Rightarrow$ eliminar o neurônio j

Heurística para γ

$$E\left|\Delta w_{ij\,BP}\right| \approx E\left|\Delta w_{ij\ \gamma F}\right|$$

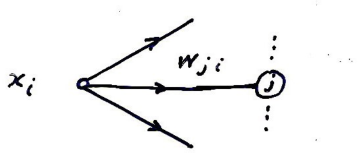
$$E\left|\Delta w_{ij\,BP}\right| \approx 2\alpha\gamma \frac{a^2}{\left(a^2 + \sum_{\forall i} w_{ij}^2\right)^2} E\left|w_{ij}\right| = \frac{\alpha\gamma}{2a}$$

$$\gamma \approx \frac{2a}{\alpha} E\left|\Delta w_{ij\,BP}\right|$$

Reduzindo entradas (alternativa à relevância)

Similar à eliminação de neurônios

$$F = F_0 + \gamma F_1 = F_0 + \gamma \sum_{\forall j} \frac{\sum_{\forall i} w_{ij}^2}{a^2 + \sum_{\forall i} w_{ij}^2}$$



$\sum_{\forall j}$ somatório para todas as entradas j

$\sum_{\forall i} w_{ij}^2$ somatório para todas as sinapses da entrada j

Poda e saturação para criação de regras lógicas

Íris

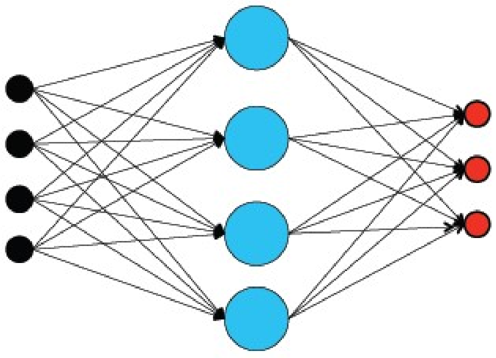
Setosa

Versicolor

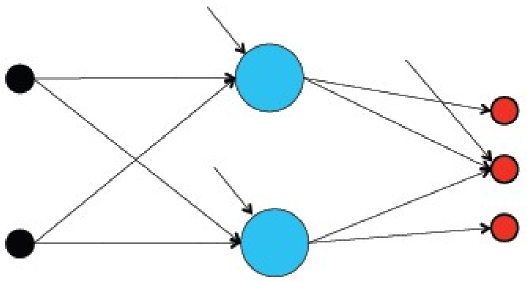
Virgínica

Entradas:	Saídas
largura da sépala	Íris setosa
comprimento da sépala	Íris versicolor
largura da pétala	Íris virgínica
comprimento da pétala	

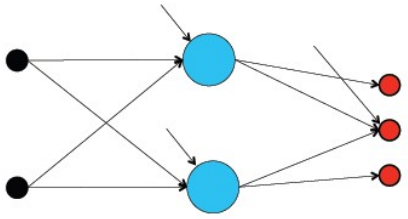
A rede inicial e a rede podada e saturada



rede inicial



rede podada e saturada



regras lógicas baseadas
diretamente nas sinapses:

1. Se 1,13 largura pétala + 2,88 comprimento pétala < 5,279 → setosa

2. Se: 1,29 largura pétala + 1,12 comprimento pétala < 10,36 → virgínica

3. Versicolor em outros casos.

regra lógica refeita (denormalizada e aproximada):

1. Se: largura da pétala + comprimento da pétala < 5 → setosa

2. Se: largura da pétala + comprimento da pétala < 10 → virgínica

3. Versicolor em outros casos.