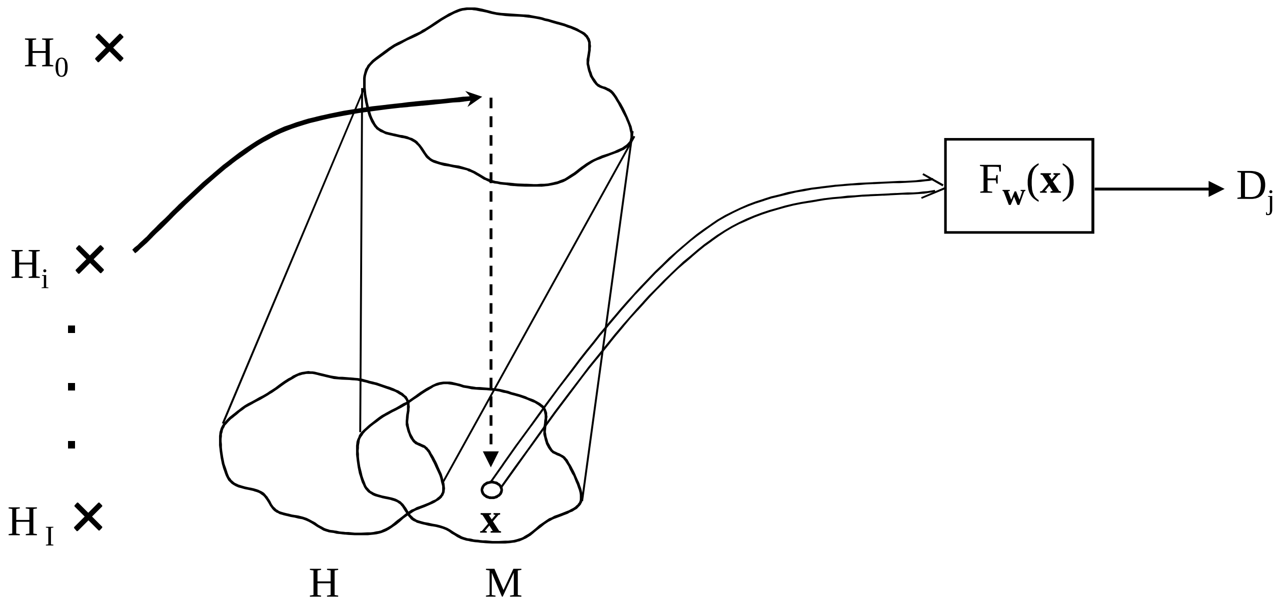
# REDES NEURONALES CON PESOS FUNCIONALES GENERADOS MEDIANTE PUERTA

# (GG-FWNN)

Aníbal R. Figueiras-Vidal
G2PI-DTSC-UCIIIM

# ÍNDICE

# 1. DECISION MÁQUINA




Máquina: información $\left\{\begin{matrix} \mathbf{x}^{(k)}, t^{(k)} \\ \mathbf{x}^{(k)}, \left(t^{(k)}\right) \end{matrix}\right\}$ (supervisión) (semisupervisión)

Propósitos: explicativo (Occam)

predictivo ("No free lunch")

# 2. CONJUNTOS

(Occam vs. Epicuro)

Ventajas:

- potencia expresiva a menor coste de diseño
- eventualmente, mejor interpretabilidad

Familias:

- <u>comités</u> : aprendices y fusión separados
- <u>cooperativos</u> : aprendices y fusión simultáneos
  - “boosting” (y NCL)
  - MoE

# 3. COMITÉS

Aprendices: "diversidad" □ arquitectura
□ aprendizaje: datos
variables
inicialización
algoritmo
. . .

(extremo: expertos regionales) (también diseños cooperativos)

Fusión: fija (promedio) / entrenable (comb. lineal)
global (anteriores) / local (mayoría; <u>puerta</u>)
. . .

Ejs. importantes: □ "bagging" (aprendices con "booststrap")
("wagging", etc.)
□ RF (variables y datos)
(se basan en aprendices <u>inestables</u>)

(RF: <u>pierden</u> inteligibilidad; paralelas a Inteligencia Colectiva… ¡mal pensada!) (fusión puerta)

# 4. MEZCLAS DE EXPERTOS (1)

A. <u>Regresión</u>

Se supone

$$p(s \mid \mathbf{x}) = \sum_{m} g_m(\mathbf{x})\, \mathbf{G}\left(s \mid \mathbf{w}_{me}^{T}\mathbf{x}_e, v_m\right)$$

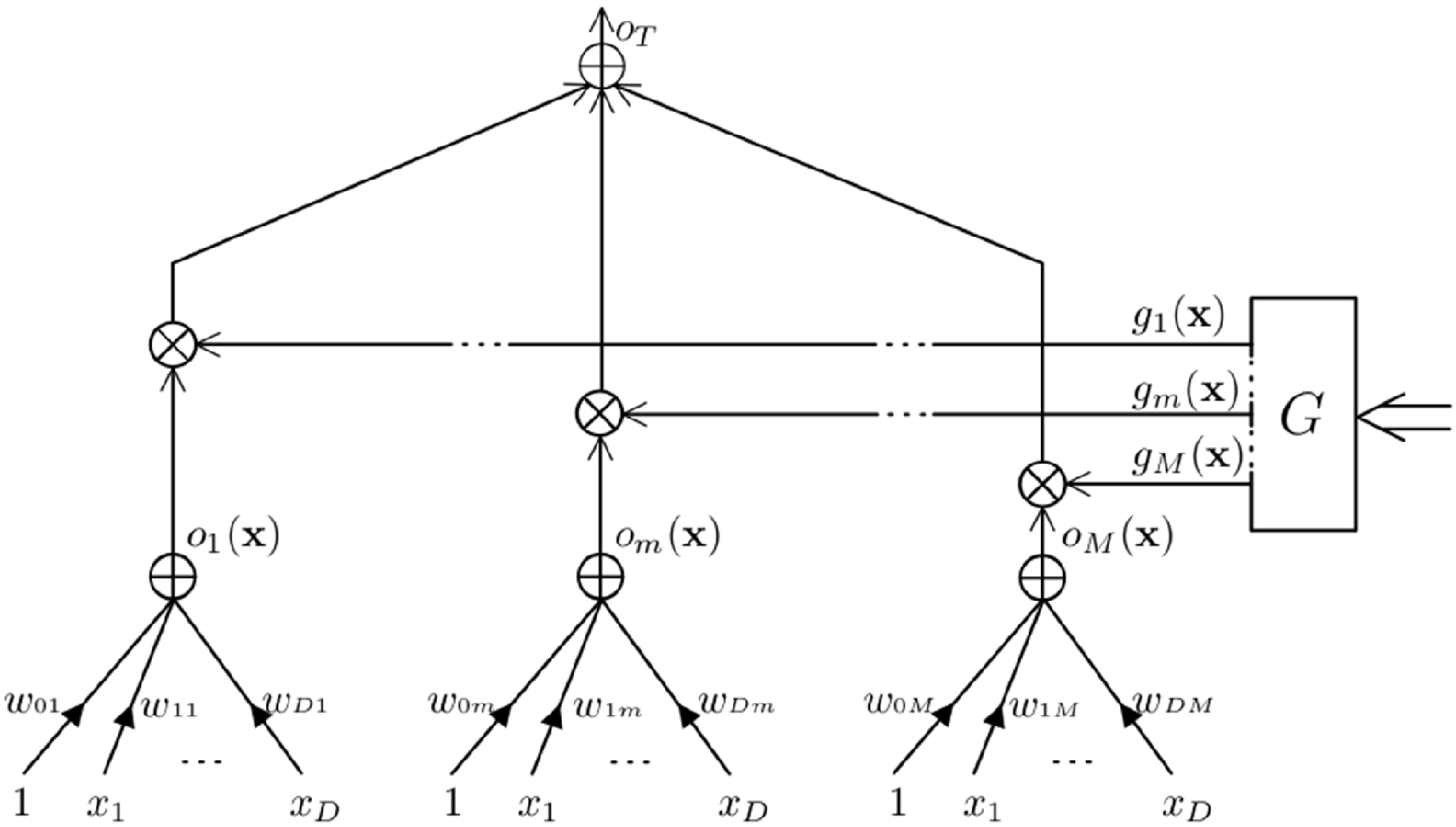
$$\Rightarrow\ E\{s \mid \mathbf{x}\} = \sum_{m} g_m(\mathbf{x})\mathbf{w}_{me}^{T}\mathbf{x}_e$$

Si $\{g_m(\mathbf{x})\}$ sencillas, entrenamiento ML directo / EM

(subir capacidad expresiva: jerárquicos)

# 4. MEZCLAS DE EXPERTOS (2)

Arquitectura



# 4. MEZCLAS DE EXPERTOS (3)

A. Entrenamiento directo

Con $g_m(\mathbf{x}) = \dfrac{\exp(\mathbf{w}_{gme}^T \mathbf{x}_e)}{\sum_{m'} \exp(\mathbf{w}_{gm'e}^T \mathbf{x}_e)} = \dfrac{\exp[y_{gm}(\mathbf{x})]}{\sum_{m'}}$

$$L(\{\mathbf{w}_{me}\},\{\mathbf{w}_{gme}\}) = \ln \sum_{m'} g_{m'} \frac{1}{\sqrt{v_{m'}}} \exp\left[-\frac{1}{2}\frac{(s-o_{m'})^2}{v_{m'}}\right] + cte$$

$$\frac{\partial L}{\partial o_m} = \frac{g_m \dfrac{1}{\sqrt{v_m}} \exp\left[-\dfrac{1}{2}\dfrac{(s-o_m)^2}{v_m}\right]}{\sum_{m'}} \frac{s-o_m}{v_m} = h_m(\mathbf{x}) \frac{s-o_m}{v_m}$$

( $h_m$ : “pertenencia” a $m$ )

# 4. MEZCLAS DE EXPERTOS (4)

$$\frac{\partial L}{\partial \mathbf{w}_{me}} = \frac{\partial L}{\partial o_m}\frac{\partial o_m}{\partial \mathbf{w}_{me}} = -h_m \frac{s - o_m}{v_m}\mathbf{x}_e \ , \quad \text{y LMS}$$

$$L = \ln \sum_{m'} \left(\exp y_{gm'}\right) \frac{1}{\sqrt{v_{m'}}} \exp\left[-\frac{1}{2}\frac{(s-o_{m'})^2}{v_{m'}}\right] - \ln \sum_{m'} \exp(y_{gm'})$$

$$\frac{\partial L}{\partial y_{gm}} = \frac{\exp(y_{gm}) \frac{1}{\sqrt{v_m}} \exp\left[-\frac{1}{2}\frac{(s-o_m)^2}{v_m}\right]}{\sum_{m'}} - \frac{\exp\left(y_{gm}\right)}{\sum_{m'}} = h_m - g_m$$

$$\frac{\partial L}{\partial \mathbf{w}_{gme}} = \frac{\partial L}{\partial y_{gm}}\frac{\partial y_{gm}}{\partial \mathbf{w}_{gme}} = (h_m - g_m)\mathbf{x}_e \ , \quad \text{y LMS}$$

## 4. MEZCLAS DE EXPERTOS (5)

En cuanto a $v_m$,

$$\frac{\partial L}{\partial v_m} = \frac{g_m \left\{ -\frac{1}{2} v_m^{-3/2} \exp\left[ -\frac{1}{2} \frac{(s-o_m)^2}{v_m} \right] + \frac{1}{\sqrt{v_m}} \exp\left[ -\frac{1}{2} \frac{(s-o_m)^2}{v_m} \right] \frac{1}{2} \frac{(s-o_m)^2}{v_m^2} \right\}}{\sum_{m'}} =$$

$$= h_m \frac{1}{2 v_m^{3/2}} \left[ \frac{(s-o_m)^2}{v_m} - 1 \right]$$ , y LMS o actualización directa

También se puede utilizar EM.

Se generaliza fácilmente para jerárquicos.

# 4. MEZCLAS DE EXPERTOS (6)

## B. Entrenamiento EM

### B.1. Versión general

Se introducen los indicadores $\{z_m(\mathbf{x})\}$ y la ddp completa

$$p(s,\mathbf{z}|\mathbf{x}) = \prod_m \left[ g_m(\mathbf{x}) p_m(s|\mathbf{x}) \right]^{z_m(\mathbf{x})}$$

La log L es

$$L(\mathbf{w}_{tot}, \mathbf{z}|\mathbf{x}_e) = \sum_m z_m(\mathbf{x}) \left[ \ln g_m(\mathbf{x}) + \ln p_m(s|\mathbf{x}) \right]$$

muestralmente

$$L_K\left(\mathbf{w}_{tot}, \{\mathbf{z}^{(n)}\}\right) = \sum_n \sum_m z_m^{(n)} \left( \ln g_m^{(n)} + \ln p_m^{(n)} \right)$$

# 4. MEZCLAS DE EXPERTOS (7)

* Paso E:

$$Q(\mathbf{w}_{tot}, \mathbf{w}_{tot}[r]) = E_{\mathbf{z}}\left\{L_K\left(\mathbf{w}_{tot}, \{\mathbf{z}^{(n)}\} \middle| \mathbf{w}_{tot}[r]\right)\right\} =$$

$$= \sum_n \sum_m E\left\{z_m^{(n)} \middle| s^{(n)}, \mathbf{x}^{(n)}, \mathbf{w}_{tot}[r]\right\} \left(\ln g_m^{(n)} + \ln p_m^{(n)}\right)$$

y como

$$E\{\;\} = \Pr\left\{z_m^{(n)} = 1 \middle| s^{(n)}, \mathbf{x}^{(n)}, \mathbf{w}_{tot}[r]\right\} = \frac{p\left(z_m^{(n)} = 1, s^{(n)} \middle| \mathbf{x}^{(n)}, \mathbf{w}_{tot}[r]\right)}{p\left(s^{(n)} \middle| \mathbf{x}^{(n)}, \mathbf{w}_{tot}[r]\right)} =$$

$$= \frac{p\left(s^{(n)} \middle| z_m^{(n)} = 1, \mathbf{x}^{(n)}, \mathbf{w}_{tot}[r]\right) \Pr\left(z_m^{(n)} = 1 \middle| \mathbf{x}^{(n)}, \mathbf{w}_{tot}[r]\right)}{p\left(s^{(n)} \middle| \mathbf{x}^{(n)}, \mathbf{w}_{tot}[r]\right)} =$$

$$= \frac{p_m\left(s^{(n)} \middle| \mathbf{x}^{(n)}, \mathbf{w}_{tot}[r]\right) g_m^{(n)}}{\sum_{m'} p_{m'}\left(s^{(n)} \middle| \mathbf{x}^{(n)}, \mathbf{w}_{tot}[r]\right) g_{m'}^{(n)}} = h_m^{(n)}(\mathbf{w}_{tot}[r])$$

resulta

$$Q(\mathbf{w}_{tot}, \mathbf{w}_{tot}[r]) = \sum_n \sum_m h_m^n(\mathbf{w}_{tot}[r]) \left(\ln g_m^{(n)} + \ln p_m^{(n)}\right)$$

# 4. MEZCLAS DE EXPERTOS (8)

* Paso M

$$\mathbf{w}_{tot}[r+1] = \arg\max_{\mathbf{w}_{tot}} Q\left(\mathbf{w}_{tot}, \mathbf{w}_{tot}[r]\right) =$$

$$= \arg\max_{\mathbf{w}_{tot}} \sum_n \sum_m h_m^{(n)}[r] \left(\ln g_m^n + \ln p_m^{(n)}\right)$$

separable en dos grupos:

$$\left\{\mathbf{w}_{gme}[r+1]\right\} = \arg\max_{\mathbf{w}_{gme}} \sum_n \sum_m h_m^{(n)}[r] \left(\mathbf{w}_{me}[r], \mathbf{w}_{gme}[r]\right) \ln g_m^{(n)}\left(\left\{\mathbf{w}_{gme}\right\}\right),\ 1 \le m \le M$$

y las ecuaciones desacopladas:

$$\left\{\mathbf{w}_{me}[r+1]\right\} = \arg\max_{\mathbf{w}_{me}} \sum_n h_m^{(n)}\left(\mathbf{w}_{me}[r], \mathbf{w}_{gme}[r]\right) \ln p_m\left(s^{(n)} \middle| x^{(n)}, \mathbf{w}_{me}\right),\ 1 \le m \le M$$

sobre lo que hay que aplicar búsqueda (gradiente estocástico: GME)

# 4. MEZCLAS DE EXPERTOS (9)

B.2. Caso gaussiano

Si $\{p_m\}$ son gaussianas

$$\partial \ln \exp\left[-\left(s - \mathbf{w}_{me}^T \mathbf{x}_e\right)^2 / 2v_m\right] / \partial \mathbf{w}_{me} = -\frac{s - \mathbf{w}_{me}^T \mathbf{x}_e}{v_m} \mathbf{x}_e$$

y para las ecuaciones del segundo grupo se tiene

$$\sum_n h_m^{(n)} [r] \left(s^n - \mathbf{w}_{me}^T \mathbf{x}_e^{(n)}\right) \mathbf{x}_e^{(n)} = \mathbf{0} \quad , \quad 1 \le m \le M$$

sistema de D+1 ecs. con D+1 incógnitas.

$\{v_m\}$ se pueden ajustar según lo ya visto.

Para las ecuaciones del primer grupo, como

$$\frac{\partial \ln g_{m'}}{\partial \mathbf{w}_{gme}} = \frac{\partial \left( \ln \exp o_{gm'} - \ln \sum_{m''} \exp\left(o_{gm''}\right) \right)}{\partial \mathbf{w}_{gme}} = \left( \delta_{mm'} - \frac{\exp o_m}{\sum_{m''} \exp o_{m''}} \right) \mathbf{x}_e = \left(\delta_{mm'} - g_m\right) \mathbf{x}_e$$

# 4. MEZCLAS DE EXPERTOS (10)

se tiene

$$\sum_{n}\sum_{m} h_{m'}^{(n)}[r]\left(\delta_{mm'} - g_{m}^{(n)}\right)\mathbf{x}_{e}^{(n)} = \mathbf{0} \;,\;\; 1 \leq m \leq M$$

$$\sum_{n}\left(h_{m}^{(n)}[r] - g_{m}^{(n)}\sum_{m'} h_{m'}^{(n)}[r]\right)\mathbf{x}_{e}^{(n)} = \mathbf{0} \;,\;\; 1 \leq m \leq M$$

$$\sum_{n}\left(h_{m}^{(n)}[r] - g_{m}^{(n)}\right)\mathbf{x}_{e}^{(n)} = \mathbf{0} \;,\;\; 1 \leq m \leq M$$

siendo $g_{m}^{(n)} = g_{m}^{(n)}\left(\left\{\mathbf{w}_{gme}\right\}\right)$;

se trata de un sistema de M ecuaciones no lineales acopladas en $\left\{\mathbf{w}_{gme}\right\}$

# 4. MEZCLAS DE EXPERTOS (11)

C. Caso de clasificación

C.1. Método de Newton

Si $\mathrm{F}$ es cuadrática

$$\mathrm{F}(\mathbf{w}) = \mathrm{F}(\mathbf{w}_0) + \left.\frac{\partial \mathrm{F}}{\partial \mathbf{w}}\right|_0 (\mathbf{w} - \mathbf{w}_0) + \frac{1}{2}(\mathbf{w} - \mathbf{w}_0)^{\mathrm{T}} \left.\frac{\partial^2 \mathrm{F}}{\partial \mathbf{w} \partial \mathbf{w}^{\mathrm{T}}}\right|_0 (\mathbf{w} - \mathbf{w}_0)$$

$$\left( \left.\frac{\partial \mathrm{F}}{\partial \mathbf{w}}\right|_0 = \mathrm{grad}\ \mathrm{F}\big|_0 ; \left.\frac{\partial^2 \mathrm{F}}{\partial \mathbf{w} \partial \mathbf{w}^{\mathrm{T}}}\right|_0 = \mathrm{H}_{\mathrm{F}}\big|_0 \right)$$

# 4. MEZCLAS DE EXPERTOS (12)

y si $\mathbf{w}^*$ es el max/min

$$\left.\frac{\partial F}{\partial \mathbf{w}}\right|_* = \mathbf{0} = \left.\frac{\partial F}{\partial \mathbf{w}}\right|_0 + \left.\frac{\partial^2 F}{\partial \mathbf{w} \partial \mathbf{w}^T}\right|_0 (\mathbf{w}^* - \mathbf{w}_0)$$

de donde

$$\mathbf{w}^* = \mathbf{w}_0 - \left[ \left.\frac{\partial^2 F}{\partial \mathbf{w} \partial \mathbf{w}^T}\right|_0 \right]^{-1} \left.\frac{\partial F}{\partial \mathbf{w}}\right|_0$$

y se llega al max/min en un paso.

Si no es cuadrático pero es suave, y $\mathbf{w}_0$ es cercano a $\mathbf{w}^*$, la igualdad es aproximada y basta con iterar.

# 4. MEZCLAS DE EXPERTOS (13)

C.2. Algoritmo IRLS ("Iterative Reweighted Least Squares")

Si se trata de un objetivo muestral con la forma

$$F = \sum_n F\left(\mathbf{w}_e^T \mathbf{x}_e^{(n)}\right):$$

$$\frac{\partial F}{\partial \mathbf{w}_e} = \sum_n F'\left(\mathbf{w}_e^T \mathbf{x}_e^{(n)}\right)\mathbf{x}_e^{(n)} = \mathbf{X}_e^T \boldsymbol{\Omega} \mathbf{e} \qquad \left((D+1)\times 1\right)$$

$$\frac{\partial^2 F}{\partial \mathbf{w}_e \partial \mathbf{w}_e^T} = \sum_n F''\left(\mathbf{w}_e^T \mathbf{x}_e^{(n)}\right) \mathbf{x}_e^{(n)} \mathbf{x}_e^{(n)T} = -\mathbf{X}_e^T \Omega\, \mathbf{X}_e \qquad \left((D+1)\times(D+1)\right)$$

# 4. MEZCLAS DE EXPERTOS (14)

con

$$\mathbf{X}_e = \begin{bmatrix} 1 & x_1^{(1)} & \cdots & x_D^{(1)} \\ 1 & x_1^{(2)} & \cdots & x_D^{(2)} \\ \vdots & \vdots & & \vdots \\ 1 & x_1^{(N)} & \cdots & x_D^{(N)} \end{bmatrix} \qquad \left(N \times (D+1)\right)$$

$$\Omega = -\begin{bmatrix} F''(\mathbf{w}_e^T \mathbf{x}_e^{(1)}) & & 0 \\ & \ddots & \\ 0 & & F''(\mathbf{w}_e^T \mathbf{x}_e^{(N)}) \end{bmatrix} \qquad \left(N \times N\right)$$

$$\mathbf{e} = -\begin{bmatrix} \dfrac{F'(\mathbf{w}_e^T \mathbf{x}_e^{(1)})}{F''(\mathbf{w}_e^T \mathbf{x}_e^{(1)})} \\ \vdots \\ \dfrac{F'(\mathbf{w}_e^T \mathbf{x}_e^{(N)})}{F''(\mathbf{w}_e^T \mathbf{x}_e^{(N)})} \end{bmatrix} \qquad \left(N \times 1\right)$$

# 4. MEZCLAS DE EXPERTOS (15)

y las iteraciones resultan

$$\begin{aligned}
\mathbf{w}_e[r+1] &= \mathbf{w}_e[r] + \left(\mathbf{X}_e^T\,\Omega[r]\,\mathbf{X}_e\right)^{-1}\mathbf{X}_e^T\,\Omega[r]\,\mathbf{e}[r] = \\
&= \left(\mathbf{X}_e^T\,\Omega[r]\,\mathbf{X}_e\right)^{-1}\mathbf{X}_e^T\,\Omega[r]\,\mathbf{X}_e\,\mathbf{w}_e[r] + \left(\mathbf{X}_e^T\,\Omega[r]\,\mathbf{X}_e\right)^{-1}\mathbf{X}_e^T\,\Omega[r]\,\mathbf{e}[r] = \\
&= \left(\mathbf{X}_e^T\,\Omega[r]\,\mathbf{X}_e\right)^{-1}\mathbf{X}_e^T\,\Omega[r]\left\{\mathbf{X}_e\mathbf{w}_e[r] + \mathbf{e}[r]\right\} = \\
&= \left(\mathbf{X}_e^T\,\Omega[r]\,\mathbf{X}_e\right)^{-1}\mathbf{X}_e^T\,\Omega[r]\,\mathbf{v}[r]
\end{aligned}$$

que es el algoritmo IRLS.

# 4. MEZCLAS DE EXPERTOS (16)

C.3. Forma exponencial del modelo de Bernoulli

Para una s binaria, el modelo mezcla de Bernoulli es

$$p\left(s|\mathbf{x}\right)=\sum_{m} g_m(\mathbf{x})o_m^s(\mathbf{x})\left[1-o_m(\mathbf{x})\right]^{1-s},0\le o_m(\mathbf{x})\le 1\ \ \left(\text{prob. de }1\right)$$

escribiendo

$$o_m^s(1-o_m)^{1-s}=\exp\left\{\ln\left[o_m^s(1-o_m)^{1-s}\right]\right\}=$$

$$=\exp\left[s\ln\frac{o_m}{1-o_m}+\ln(1-o_m)\right]=\exp\left[s\eta_m-\ln(1+\exp\eta_m)\right]$$

con $\eta_m=\ln\dfrac{o_m}{1-o_m}\Rightarrow o_m=\dfrac{\exp\eta_m}{1+\exp\eta_m}$

# 4. MEZCLAS DE EXPERTOS (17)

Empleando como aprendices decisores softmax lineales

$$o_m = \mathrm{sgm}\left(\mathbf{w}_{me}^T \mathbf{x}_e\right) = \frac{\exp\left(\mathbf{w}_{me}^T \mathbf{x}_e\right)}{1 + \exp\left(\mathbf{w}_{me}^T \mathbf{x}_e\right)}$$

resulta

$$\eta_m = \ln \frac{\exp\left(\mathbf{w}_{me}^T \mathbf{x}_e\right) / \left[1 + \exp\left(\mathbf{w}_{me}^T \mathbf{x}_e\right)\right]}{1 - \exp\left(\mathbf{w}_{me}^T \mathbf{x}_e\right) / \left[1 + \exp\left(\mathbf{w}_{me}^T \mathbf{x}_e\right)\right]} = \ln\left[\exp\left(\mathbf{w}_{me}^T \mathbf{x}_e\right)\right] = \mathbf{w}_{me}^T \mathbf{x}_e$$

y:

$$p\left(s|\mathbf{x}\right) = \sum_m g_m(\mathbf{x}) \exp\left\{s\, \mathbf{w}_{me}^T \mathbf{x}_e - \ln\left[1 + \exp\left(\mathbf{w}_{me}^T \mathbf{x}_e\right)\right]\right\}$$

# 4. MEZCLAS DE EXPERTOS (18)

C.4. Aplicación

Yendo a las ecuaciones desacopladas para las $\{\mathbf{w}_{me}\}$ en el EM

$$\{\mathbf{w}_{me}[r+1]\} = \arg\max_{\mathbf{w}_{me}} \sum_n h_m^{(n)}[r] \ln p_m\left(s^{(n)} \middle| \mathbf{x}^{(n)}, \mathbf{w}_{me}\right)$$

basta introducir en el IRLS para cada m la ponderación $h_m^{(n)}[r]$ :

$$F = h_m[r]\left\{s\,\mathbf{w}_{me}^T\mathbf{x}_e - \ln\left[1+\exp\left(\mathbf{w}_{me}^T\mathbf{x}_e\right)\right]\right\}$$

$$F' = h_m[r]\left[s - \frac{\exp\left(\mathbf{w}_{me}^T\mathbf{x}_e\right)}{1+\exp\left(\mathbf{w}_{me}^T\mathbf{x}_e\right)}\right] = h_m[r]\left[s - \mathrm{sgm}\left(\mathbf{w}_{me}^T\mathbf{x}_e\right)\right]$$

$$F'' = -h_m\,\mathrm{sgm}\left(\mathbf{w}_{me}^T\mathbf{x}_e\right)\left[1-\mathrm{sgm}\left(\mathbf{w}_{me}^T\mathbf{x}_e\right)\right]$$

# 4. MEZCLAS DE EXPERTOS (19)

con lo que se tiene:

$$e_{nm} = \frac{s^{(n)} - \mathrm{sgm}\left(\mathbf{w}_{me}^{T}\mathbf{x}_{e}^{(n)}\right)}{\mathrm{sgm}\left(\mathbf{w}_{me}^{T}\mathbf{x}_{e}^{(n)}\right)\left[1-\mathrm{sgm}\left(\mathbf{w}_{me}^{T}\mathbf{x}_{e}\right)\right]}$$

$$\Omega_{nm} = h_{m}^{(n)}[\,r\,]\,\mathrm{sgm}\left(\mathbf{w}_{me}^{T}\mathbf{x}_{e}^{(n)}\right)\left[1-\mathrm{sgm}\left(\mathbf{w}_{me}^{T}\mathbf{x}_{e}^{(n)}\right)\right]$$

y se puede aplicar cómodamente el IRLS.

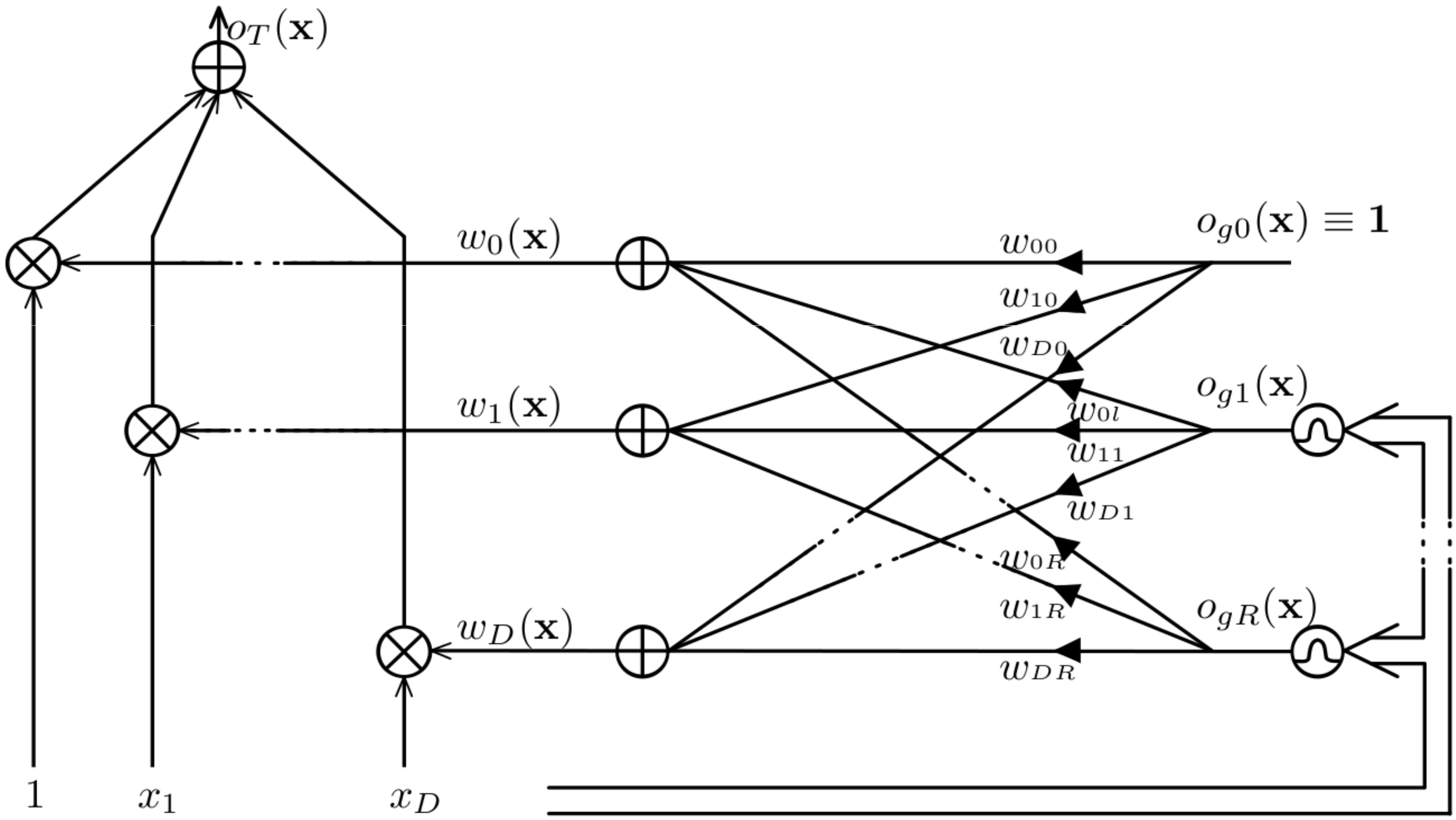
# 5. GG-FWNN (1)

Ecuaciones MoE (regresión)

$$
\begin{aligned}
o_T(\mathbf{x}) &= \sum_m g_m(\mathbf{x}) o_m(\mathbf{x}) = \\
&= \sum_m g_m(\mathbf{x}) \sum_d w_{dm} x_d = \\
&= \sum_d \left( \sum_m w_{dm} g_m(\mathbf{x}) \right) x_d = \\
&= \sum_d w_d(\mathbf{x}) x_d
\end{aligned}
$$

# 5. GG-FWNN (2)

Arquitectura



# 5. GG-FWNN (3)

Reformulación:

$$o_T(\mathbf{x}) = \sum_{d=0}^{D} w_d(\mathbf{x}) x_d =$$

$$= \sum_{d=0}^{D} \sum_{r=0}^{R} w_{dr} o_{gr}(\mathbf{x}) x_d =$$

$$= w_{00} + \sum_{r=m}^{R} w_{0r} o_{gr}(\mathbf{x}) + \sum_{d=m}^{D} \sum_{r=0}^{R} w_{dr} o_{gr}(\mathbf{x}) x_d =$$

$$= \sum_{s=0}^{S} a_s z_s(\mathbf{x})$$

$$z_s(\mathbf{x}) = o_{gr}(\mathbf{x}) x_d$$

permite MM con $\{z_s(\mathbf{x})\}$

# 5. GG-FWNN(4)

Selección de centroides

A. Preselección (Método de Shin-Choi)

Se aplica un k-NN, con $k$ el primer valor que exceda $2N\hat{P}_e$, siendo $\hat{P}_e$ la tasa de error con un $1-NN$, y se estima la probabilidad de que $\mathbf{x}^{(n)}$ pertenezca a la clase:

$$P_j\left(\mathbf{x}^{(n)}\right)=\frac{k_j^{(n)}}{k}$$

y, a partir de ese valor, se definen

- la proximidad a la frontera

$$pf\left(\mathbf{x}^{(n)}\right)=-\sum_{j=-1,1} P_j\left(\mathbf{x}^{(n)}\right)\log_2\left[P_j\left(\mathbf{x}^{(n)}\right)\right]$$

- la corrección

$$c\left(\mathbf{x}^{(n)}\right)=P_{d^{(n)}}\left(\mathbf{x}^{(n)}\right)$$

y se preseleccionan las muestras con $pf>0$ y $c>1/2$

# 5. GG-FWNN(5)

B. Selección (Método APC-III de Hwang-Bang)

- se selecciona la muestra de mayor $pf$ ,
- se excluyen las muestras en una hiperesfera de radio $h$ ,
- si quedan muestras, se itera

( $h$ es un parámetro de diseño).

# 6. GG-FWNN(6)

Prestaciones (10x10-fold) :

| | SVM | RAB | GG-FWNN 1 | GG-FWNN 2 | GG-FWNN 3 |
|---|---|---|---|---|---|
| aba | 19.8 | 19.4±0.02 | 19.8±0.2 | 19.6 | **18.9** |
| bre | 3.2 | 2.6±0.5 | 2.9±0 | **2.1** | 3.2 |
| con | 29.3±1.4 | 29.0±0.2 | **28.6±1.1** | 29.7±0.8 | 28.8±1.2 |
| cra | 3.8 | 2.5±0 | 3.7±0 | 1.2 | **0** |
| hep | 14.5 | 8.6±1.6 | 13.8±2.3 | **6.4** | 14.5 |
| ima | 3.2 | **2.5±0.04** | 3.8±0.4 | 5.0 | 2.6 |
| ion | **2.0** | 4.9±0.9 | 3.6±0.8 | 5.3 | 5.3 |
| kwo | 12.1 | **11.7±0.01** | 11.8±0.06 | 12.4 | 12.1 |
| pho | **11.1±0.4** | 14.0±0.07 | 11.9±0.4 | 14.7±0.4 | 11.8±0.5 |
| rip | 9.6 | 9.7±0.01 | 9.6±0 | 9.6 | **9.1** |
| spa | 6.3±0.7 | **5.9±0.09** | 6.3±0.6 | 6.3±0.6 | 6.2±0.7 |
| tic | 1.6±0.5 | **0.8±0.19** | **0.8±06** | 26±0.8 | 22.9±4.3 |

# 5. GG - FWNN (7)

Prestaciones (25x10-fold)

| | SVM | RAB | GG-FWNN 1 | GG-FWNN 2 | GG-FWNN 3 |
|---|---|---|---|---|---|
| hep | 14.5 | 8.6±1.6 | 8.2±1.0 | **6.4** | 12.9 |
| ion | **2.0** | 4.9±0.9 | 5.1±0.9 | **2.0** | 6.7 |
| tic | 1.7±0.3 | 0.8±0.19 | **0.4±0.5** | 1.7±0.3 | 1.7±0.3 |

# 6. MEJORAS Y EXTENSIONES

* Mejoras mediante técnicas complementarias
  - otras selecciones de centroides
  - ponderación de muestras
  - reducción / construcción de variables

* Mejoras mediante modificaciones
  - selección de $\{z_s\}$
  - versiones adaptativas “on line”
  - otras técnicas constructivas

* Extensiones
  - regresión
  - FWNN directas

# 7. UNA NOTA SOBRE INTEGRACIÓN H-M (1)

∗ Es obvio que conviene reducir la reluctancia de los expertos

- socialmente
- sicológicamente: observar como un niño
  evaluar como un sabio
  actuar como un artista

∗ También es evidente que hay que mejorar la interpretabilidad de las máquinas

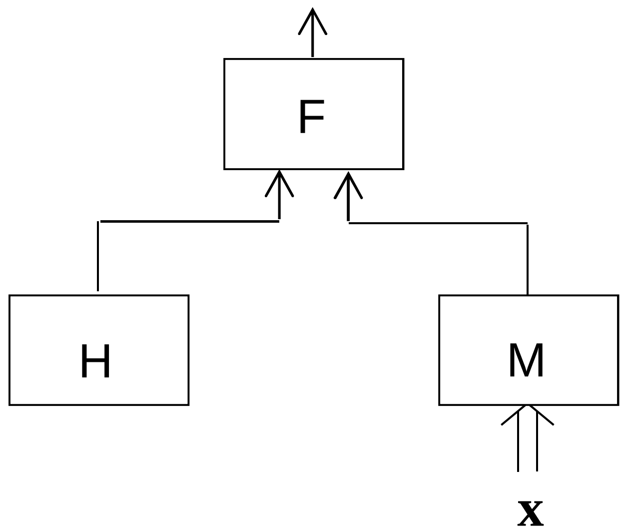
∗ Pero el Gran Desafío es la <u>integración ≷ coevolución</u>

(requiere: sicólogos, sociólogos, matemáticos, ingenieros… en equipo;
también mucho trabajo teórico y experimental).

# 7. UNA NOTA SOBRE INTEGRACIÓN H-M (2)

* Primeros intentos de integración:
  - Reglas / Sistemas Expertos: dificultad de elicitación
  - Grafos para redes Bayesianas: análogo riesgo

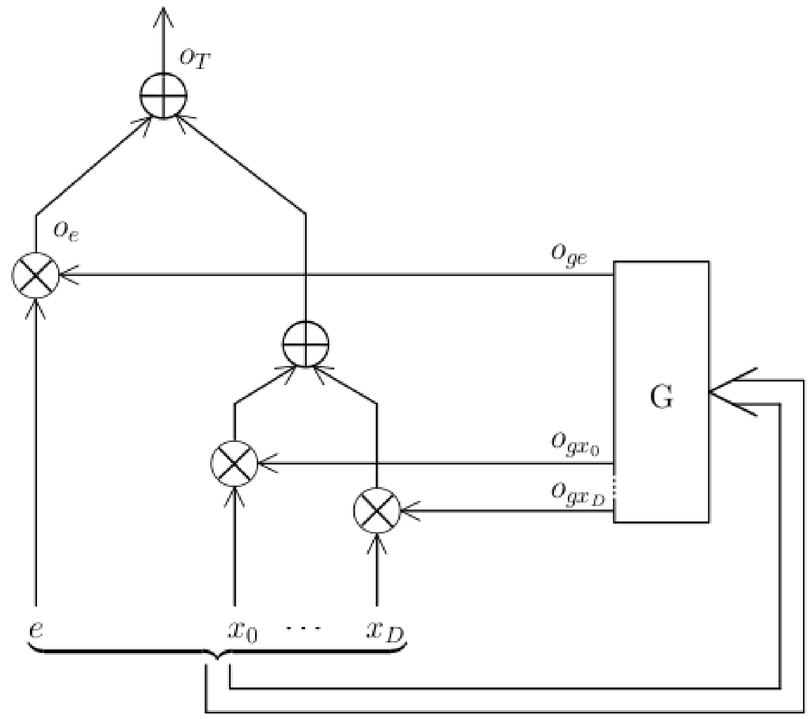
* Son “aprendices diversos” => fusión



# 7. UNA NOTA SOBRE INTEGRACIÓN H-M (3)

Requisitos
- personalización
- coevolución
  - M adaptativa
  - M interpretable

Fusión con GG-FWNN:



# 16. UNA NOTA SOBRE INTEGRACIÓN H-M (4)

También <u>Inteligencia Colectiva</u>