



INTELIGENCIA COLECTIVA Y APRENDIZAJE MÁQUINA

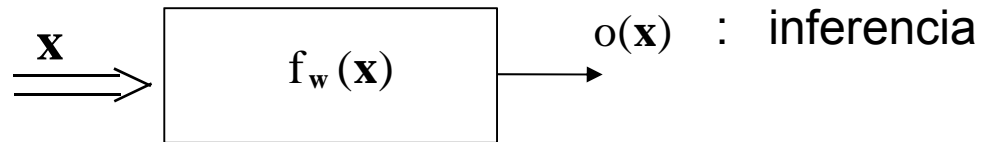
Aníbal R. Figueiras Vidal
CU, UC3M
AN, RAIng

ÍNDICE



1. Aprendizaje máquina
2. Conjuntos
3. Comités
4. Consorcios
5. Mezclas de expertos
6. “Boosting”
7. Revisión de líneas de trabajo
8. Líneas “rutinarias”
9. Líneas “incrementales”
10. Líneas “radicales”
11. Usos
12. Ámbitos de aplicación
13. Unos humildes consejos
14. Reconocimientos
15. Bibliografía

1. Aprendizaje Máquina



f : familia de funciones \approx arquitectura

w : parámetros : búsqueda

ejemplos $\{\mathbf{x}^{(k)}, d^{(k)}\}$: $\left\{ \begin{array}{l} - \text{criterio} \\ - \text{algoritmo} \end{array} \right.$

Dificultades :

* f ?

* K vs. complejidad ?

* convergencia de la búsqueda ?

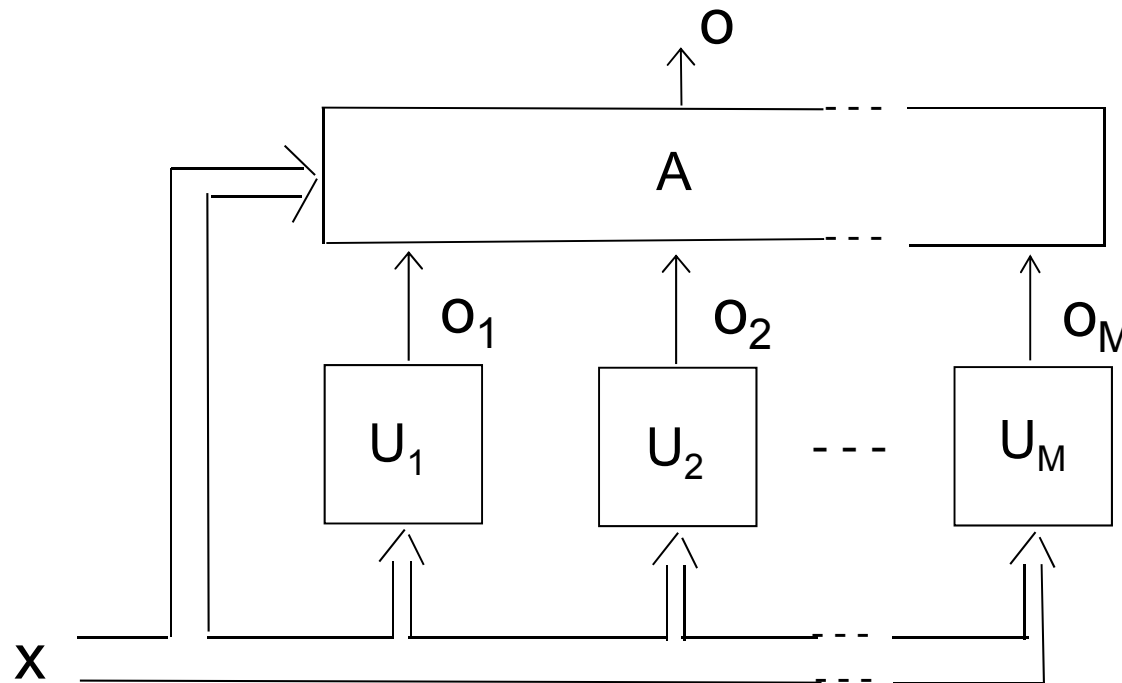
\Rightarrow como inteligencia colectiva : conjuntos

- independencia

- distribución

2. Conjuntos (1)

- Forma habitual: aprendices (unidades) + agregación



(Visión más general “(mecánica”): componentes + integración)



2. Conjuntos (2)

* Ventajas

- alivia elección f, alivia búsqueda
- mayor potencia expresiva
- incluso mejor comprensión

* Requisitos

- diversidad (complementaria { principios
(aprendices) "ad hoc": ?)
- adecuación (sencillez vs. potencia)
(agregación)
- (eventual poda)

* Posibilidades

- diseño separado → comités
- diseño conjunto → consorcios



3. Comités (1)

- * Fuentes de diversidad (convencionales)

- en los datos:
 - remuestreo (incl. generación)
 - ruido (incl. reetiquetado)
 - variables
- en la arquitectura (↓ ; pero local vs. global!)
(+ parámetros no entrenables: dim, disp, etc.)
- en la búsqueda:
 - coste
 - tipo
 - inicialización

- * Tipos de agregación (convencionales)

- selección (ej.: expertos) / combinación
- de resultados (ej.: mayoría) / de salidas (ej.: media)
- fijas (ej.: mayoría) / entrenables (ej.: comb. lineal)
- globales (ej.: comb. lineal) / locales (ej.: puerta local)



3. Comités (2)

- * Bagging (“bootstrap and aggregating”; L. Breiman)
 - aprendices iguales con remuestreos “bootstrap”
 - agregación directa (son “grandes”): media, mayoría,...

Wagging: muestras ponderadas según distribuciones
(Poisson: (bootstrap) Bagging)

- * Selvas aleatorias
 - árboles en subespacios
 - o/y con submuestreo
 - o/y ramificación probabilística
 - agregación directa
(se pierde la (falsa) “interpretabilidad”)

- * Stacking
 - procediendo por capas (¿”deep”?)



4. Consorcios

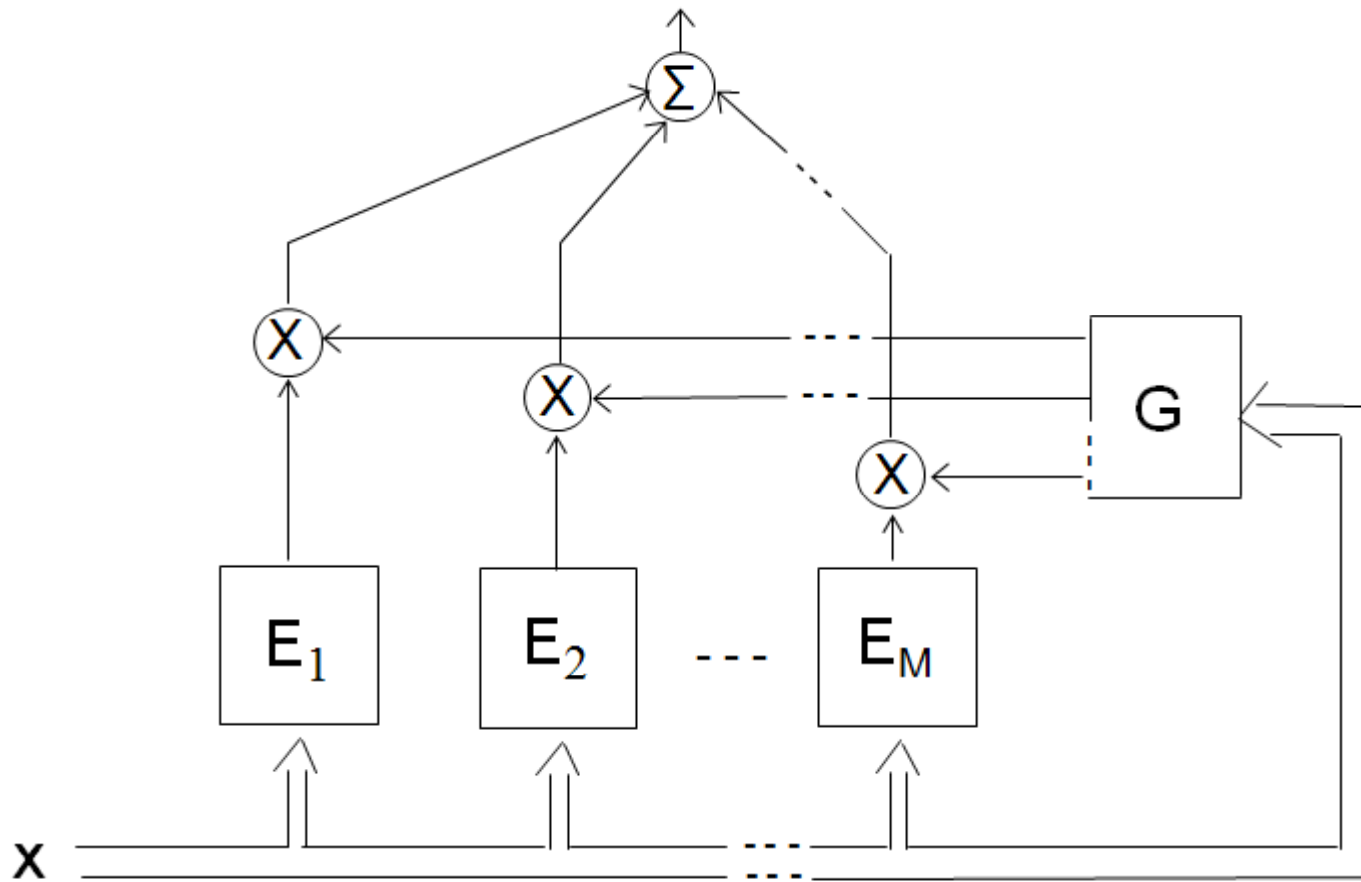
- * Se basan en mecanismos de cooperación / competición
(algunos)

- * Se construyen conjuntamente:
 - en bloque: Mezclas de Expertos
 - secuencialmente: “Boosting”

(→ mayor eficacia)

5. Mezclas de Expertos (1)

- Las unidades (expertos) compiten por subregiones
- La puerta o árbitro combina “blandamente” sus salidas (cooperación)





5. MoEs (2)

Entrenamiento

- Regresión: salida, media de GM $\sum_i \rho_i(\mathbf{x}) G(y | m_i(\mathbf{x}), v_i(\mathbf{x}))$
entrenamiento: ML (EM)

(fácil si $m_i(\mathbf{x}) = \mathbf{w}_{ei}^T \mathbf{x}_e$

$$v_i(\mathbf{x}) = v_i$$

$$\rho_i(\mathbf{x}) = \frac{\exp(\mathbf{w}_{gi}^T \mathbf{x}_e)}{\sum_{i'} \exp(\mathbf{w}_{gi'}^T \mathbf{x}_e)}$$

pero poco poder expresivo: \longrightarrow conjuntos jerárquicos)

- Decisión: análogo, con forma exponencial de dicotómica repetida
(necesita IRLS...
pero malos resultados)



5. MoEs (3)

- * Fortalezas:

- eficaces
- comprensibles

- * Debilidades:

- limitaciones estructurales (→ jerárquicas)
- mal en decisión

6. Boosting (1)

Aprendices débiles:

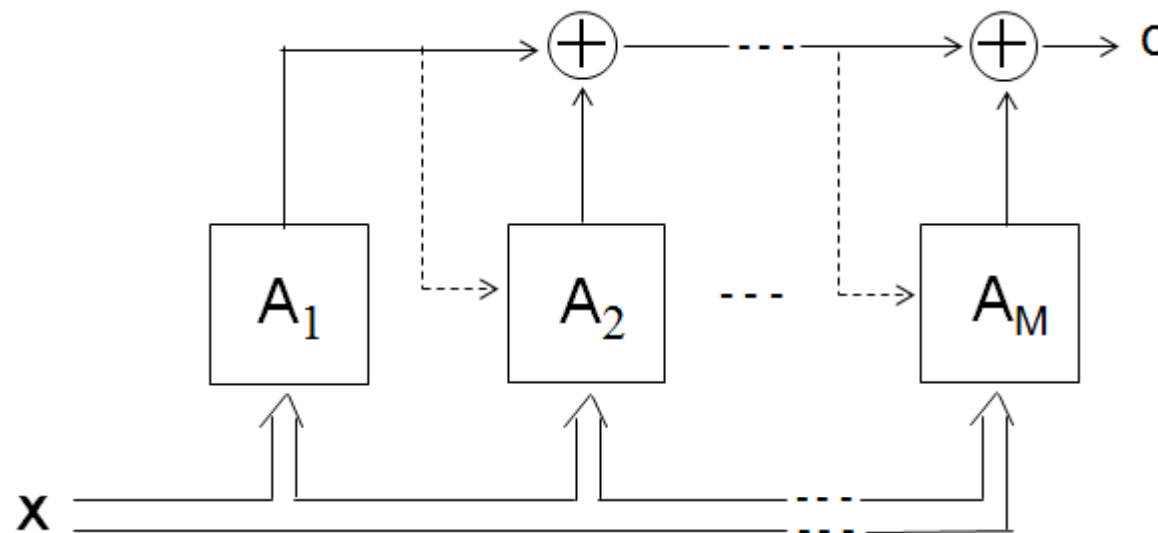
se añaden uno a uno para ayudar a los anteriores

con las muestras difíciles \longrightarrow mayor ponderación (énfasis)

(cooperación-competición: grado)

Agregación:

lineal, para minimizar coste exponencial $\exp(-df)$ (“cerrado”)





6. Boosting (2)

Originales: AdaBoost, aprendices duros
Real AdaBoost, aprendices blandos

RAB: se minimiza una cota superior del error

1. $\{p_0^{(k)}\} = 1/K$
2. $l = 1, \dots, L$
 - 2.1. $o_l = \arg \max_{o_l} \sum_k p_{l-1}^{(k)} d^{(k)} o_l^{(k)}$ (- convencional pond.)
 - 2.2.
 - a. $\gamma_l = \sum_k p_{l-1}^{(k)} d^{(k)} o_l^{(k)}$
 - b. $\alpha_l = \frac{1}{2} \ln \left(\frac{1 + \gamma_l}{1 - \gamma_l} \right)$
 - 2.3. $p_l^{(k)} = p_{l-1}^{(k)} \exp(-\alpha_l d^{(k)} o_l^{(k)}) / Z_l$
3. $o_{cl}(\mathbf{x}) = \text{sgn} \left(\sum_{l=1}^L \alpha_l o_l(\mathbf{x}) \right)$
4. Condición de parada: no \longrightarrow 2



6. Boosting (3)

- * Fortalezas:

- gran resistencia al sobreajuste (por “margen”: ?
por énfasis: sí;
L. Breiman, “Arcing”,
“Adaptive reweighting and combining”)
- excelentes prestaciones

- * Debilidades:

- no acepta aprendices estables
- no es muy eficaz con aprendices fuertes
- frente a mucho ruido u “outliers”
- frente a asimetrías

(Muchas modificaciones y variantes)



7. Revisión de líneas de trabajo

- * “Rutinarias”
- * “Incrementales”
- * “Radicales”

Veremos también:

- * Usos
- * Ámbitos de aplicación



8. Líneas “rutinarias” (1)

Las típicas extensiones

- para problemas multiclase
- para problemas multietiqueta
- para problemas “sensibles a los costes” ($C_{ji}(\mathbf{x})$)

...

y modificaciones

- otros costes
- otros énfasis

...

(O.p.: Deben ser subproductos de la necesidad;
pero no hay que despreciarlas)



8. Líneas “rutinarias” (2)

Casos (propios) de éxito (corrigiendo debilidades operativas):

A. Énfasis mixto: $\exp\left\{\lambda\left(o_1^{(k)} - d^{(k)}\right)^2 - (1 - \lambda)o_1^{(k)2}\right\}$ (ind. prox. frontera)

	ab	co	im	kw	ph	ri	sp	ti
RAB	19.38 ±0.15	29.00 ±1.45	2.46 ±0.31	11.71 ±0.05	14.04 ±0.52	9.73 ±0.09	5.94 ±0.61	<u>0.75</u> ±0.55
MRAB	18.97 ±0.13	<u>28.54</u> ±1.25	<u>2.31</u> ±0.29	11.66 ±0.05	13.43 ±0.63	9.41 ±0.18	<u>5.75</u> ±0.51	0.79 ±0.55



8. Líneas “rutinarias” (3)

B. Énfasis suavizado:

$$D_1'(\mathbf{x}^{(n)}) = \alpha D_1(\mathbf{x}^{(n)}) + \frac{1-\alpha}{K} \sum_{NN_K} D_1(\mathbf{x}^{(n')})$$

	aba	bre	cra	cre	dia	ger	hep	ima	ion	kwo	rip	wav
RAB	19.4 ±0.02	2.6 ±0.4	2.5 ±0	10.14 ±0.74	20.61 ±0.68	22.27 ±0.71	8.9 ±1.8	2.99 ±0.43	4.9 ±0.9	11.7 ±0.01	9.7 ±0.01	11.65 ±0.36
K-RAB	18.98 ±0.18	2.19 ±0.34	2.5 ±0	9.86 ±0.81	20.39 ±0.66	22.11 ±0.80	6.55 ±0.89	2.86 ±0.46	4.20 ±0.90	11.59 ±0.06	9.16 ±0.21	11.31 ±0.42

(t-test)



9. Líneas “incrementales” (1)

Procesos más elaborados:

- reestructurar
- combinar (p. ej., “diversidades”)
- ...

Casos (propios) de éxito:

A. “Gate generated functional weights”

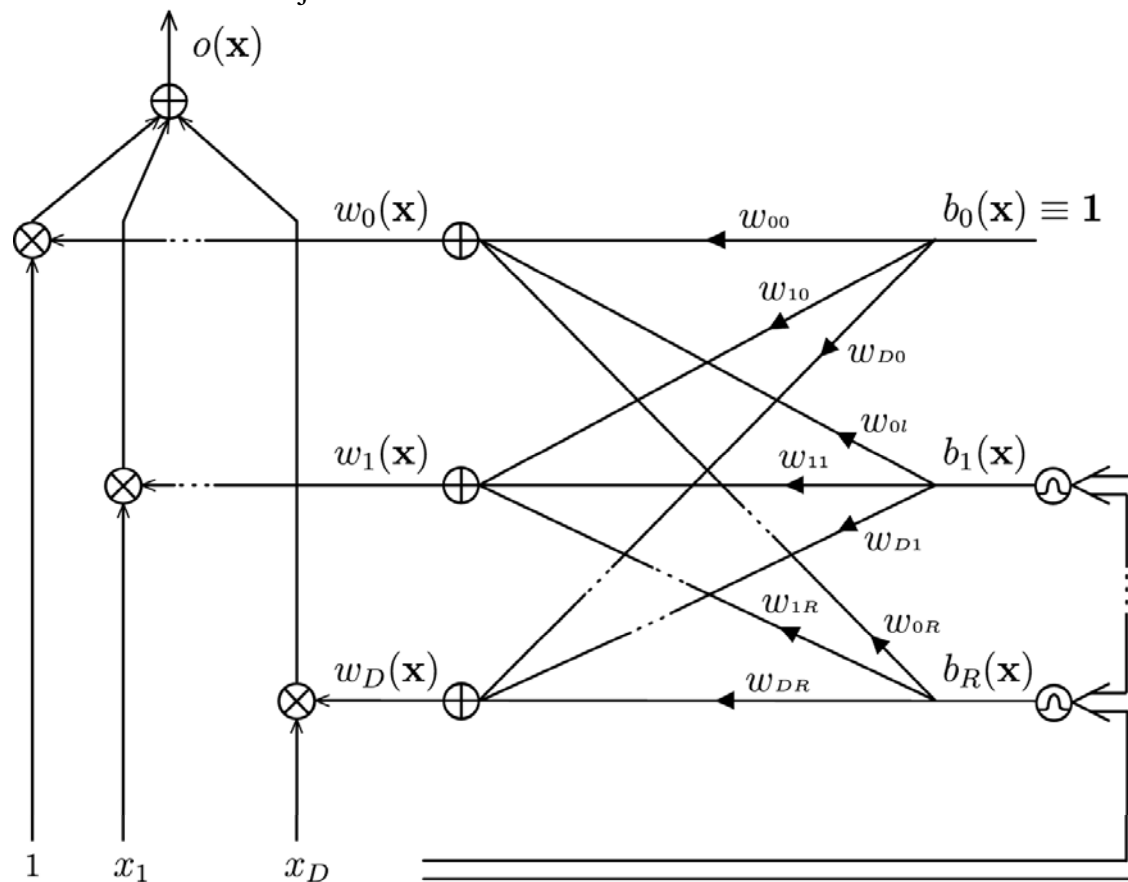
MoE:

$$\begin{aligned}o(\mathbf{x}) &= \sum_j g_j(\mathbf{x}) \sum_i w_{eji} X_i = \sum_i \left(\sum_j g_j(\mathbf{x}) w_{eji} \right) X_i = \\ &= \sum_i w_i(\mathbf{x}) X_i\end{aligned}$$

9. Líneas “incrementales” (2)

$w_i(\mathbf{x})$ generados por puerta con núcleos: $w_i(\mathbf{x}) = \sum_j w_{ij} b_j(\mathbf{x})$

$$\Rightarrow o(\mathbf{x}) = \sum_i \sum_j w_{ij} b_j(\mathbf{x}) x_j \rightarrow \sum_r w_r z_r(\mathbf{x})$$





9. Líneas “incrementaños” (3)

Con $z_r(\mathbf{x})$ dado, se puede aplicar MM

	aba	bre	con	cra	hep	ima	ion	kwo	pho	rip	spa	tic
SVM	19.8	3.2	29.3 ±1.4	3.8	14.5	3.2	2.0	12.1	11.1 ±0.4	9.6	6.3 ±0.7	1.7 ±0.3
RAB	19.4 ±0.02	2.6 ±0.5	29.0 ±0.2	2.5 ±0	8.6 ±1.6	2.5 ±0.04	4.9 ±0.9	11.7 ±0.01	14.0 ±0.07	9.7 ±0.01	5.9 ±0.09	0.8 ±0.1
GGFW	18.9 (V3)	2.1 (V2)	28.5 ±0.9 (V1)	0 (V3)	6.4 (V2)	2.6 (V3)	2.0 (V2)	12.0 ±0.01 (V1)	11.8 ±0.5 (V3)	9.1 (V3)	6.2 ±0.7 (V3)	0.5 ±0.8 (V1)

(F-score)



9. Líneas “incrementales” (4)

B. Puertas en agregación “boosting”

(añade local)

	ab	br	co	cr	he	io	kw	ri
RAB	19.4 ±0.02	2.6 ±0.4	29.0 ±0.2	2.5 ±0.0	8.9 ±1.8	4.5 ±0.9	<u>11.7</u> ±0.01	9.7 ±0.01
GRAB	19.1 ±0.3	2.2 ±0.3	27.4 ±0.8	2.5 ±0.0	8.1 ±1.9	4.0 ±1.0	11.7 ±0.2	8.5 ±0.2

(t-test)



9. Líneas “incrementales” (5)

C. Boosting con SVM

(mediante: submuestreo + L1 en dual)

	ab	br	co	di	du	io	kw	ma	ri	th	tw	wa
SVM	20.1	2.5	28.3	19.3	13.9	2.0	11.8	42.2	9.5	5.3	2.5	10.5
LPSVM	20	2.5	28.6	22.3	14.0	2.0	11.7	43.5	9.4	6.7	2.8	11.2
RAB-SLPSVM	19.1 ±0.1	2.2 ±0.2	27.9 ±0.1	18.2 ±0.3	12.0 ±1.2	2.0 ±0.7	11.5 ±0.0	40.8 ±0.6	8.8 ±0.2	4.5 ±0.8	2.5 ±0.1	10.5 ±0.1



10. Líneas “radicales” (1)

Nuevas visiones del diseño.

Ejs. propios (activos):

- * Aprendiz agregador: usar GG-FW
 - permite “mejora”
 - permite postagregación
 - permite integración H-M (proyecto AHI)
- * Diversidad en agregaciones →
- * Nuevos mecanismos de cooperación – competición:
 - basados en juegos
 - basados en colectividades
 - basados en simulación de emociones (pr. AHI)
- * Diversidad en aprendizaje profundo →
- * Modularización (visión como componentes + integración)



10. Líneas “radicales” (2)

Casos (propios) de éxito

A. Diversidad en agregaciones

Aprendices base: MLPs; agregadores: mayoría (MY), máx. margen (MM)
diversidad: bagging (B), volteo de etiqueta (V)

Resultados preliminares (“omni”)

	bre	hep	ion	kwo	rip
B+MY	1.8	6.5	3.3	11.7	8.8
V+MY	1.8	6.5	1.3	11.6	9.1
B+MM	1.4	1.6	1.3	11.7	8.5
V+MM	1.8	3.2	1.3	11.8	8.8
B+MM+B+MY	1.4	<u>1.6</u>	1.3	11.7	8.7
B+MM+V+MY	1.1	3.2	1.3	11.7	8.1
V+MM+B+MY	1.4	3.2	1.3	11.7	8.3
V+MM+V+MY	1.4	<u>1.6</u>	0.7	11.7	8.3

(No se pierde)

(Problema: estabilidad → subespacios)



10. Líneas “radicales” (3)

B. Diversidad en aprendizaje profundo

(En general: representación + clasificación)

Tipo de diversidad: bagging

Aplicación: sobre las “Deep Machines” (DM) completas

DMs: Deep Belief Networks (DBM) (Hinton)

(son “Reduced Boltzmann Machines” apiladas)

Resultados sobre MNIST (dígitos manuscritos)

	B	Arq. (nodos en capa)	CE (%)
DBN	(1)	500-500-2000-10	1.26
B-DBN	10	500-500-2000-10	0.5
B-DBN	10	500-500-200-10	0.4
B-DBN	40	500-500-2000-10	~0.3
B-DBN	40	500-500-200-10	~0.3

(Obviamente, hay otras muchas posibilidades)



11. Usos

Son muy prometedores

- Sistemas de ayuda a la decisión

(/recomendación/alarma/seguridad/...)

nuestros cerebros no son perfectos: herramientas cognitivas

- Integración humanos-máquinas (pr. AHI)

... pero tampoco inútiles: intuición

y somos sociales: agrupamiento

- Sistemas distribuidos

(hay una multitud de problemas y situaciones)



12. Ámbitos de aplicación

Nada más prometedor que: { Web
Smart Cities
Smart Homes

Hay tensiones global/local/individual

⇒ Smart Soc { no sólo conectividad + cómputo: inteligencia (“magia con varita”)
aprovechar lo local
buscar el bienestar

Implicar pensar en:

- Smart administration
- Smart web sites
- Smart business
- Smart shops
- Smart employment
- Smart energy
- Smart environment
- Smart health
- Smart entertainment
- Smart transportation
- Smart information
- ...

- * No particionar (es disfuncional: vid. “mundo real”)
- * Inteligencia colectiva + inteligencia computacional
- * Aprovechar lo local (no sólo puntos fuertes)
- * Aplicar los principios vistos (reorganizar, combinar,...)



13. Unos humildes consejos

Hay más oportunidades de las que se puede pensar en abordar ...

- * Ser ambicioso
- * Entender la ventaja general de la diversidad
y la cooperación-competición
- * Ser creativo (no sólo innovador): mirar desde arriba
predecir beneficio
- * Olvidar la soberbia, la certeza, el tema prioritario/único, ..., aprender de los errores
- * Reflexionar acerca de todo (intuitiva y analíticamente)
- * Reconocer que la realidad es un texto, y leerlo.



14. Reconocimientos

En estas líneas de trabajo son co-investigadores internos:

Dr. Emilio Ortiz García

Dr. Miguel Lázaro Gredilla

Dr. Efraín T. Mayhua López

y mis doctorandos

Adil Omari

Luis Muñoz González

Ricardo F. Alvear Sandoval

Anas Ahachad

Fernando Rabanal Presa

Juan José Choquehuanca Zevallos

Rafael Hernández Murcia

y también hay colaboración externa con

Prof. Dr. David Ríos Insua (RACEFN, URJC)

Dr. Alberto Suárez (UAM)

Dr. Lior Rokach (Ben Gurion Univ. Negev, Israel)



15. Bibliografía (1)

Si tuviera que elegir entre textos introductorios al aprendizaje máquina citarían los de Chris Bishop:

- C. M. Bishop: *Neural Networks for Pattern Recognition*. Oxford, UK: Oxford Univ. Press; 1995.
- C. M. Bishop: *Pattern Recognition and Machine Learning*. New York, NY: Springer; 2006.

En cuanto a conjuntos de máquinas, mis preferidos son

- L.I. Kuncheva: *Combining Pattern Classifiers. Methods and Algorithms*. Hoboken, NJ: Wiley; 2004.
- L. Rokach: *Pattern Classification Using Ensemble Methods*. Singapore: World Scientific; 2010.
- R.E. Shapire, Y. Freund: *Boosting. Foundations and Algorithms*. Cambridge, MA: MIT Press; 2012



15. Bibliografía (2)

En ellos se encuentran destacadas las principales referencias a Bagging, Selvas Aleatorias, Stacking, Mezclas de Expertos y Boosting (Arcing). Debo añadir, por completitud, el artículo de revisión

- S.E. Yuksel, J.N. Wilson, P.D. Gader, “Twenty years of Mixture of Experts”, IEEE Trans. Neural Networks and Learning Systems, vol. 33, pp. 1177-1193, 2012.

En cuanto a contribuciones propias, en la comunicación invitada

- A.R. Figueiras-Vidal, L. Rokach, “An exploration of research directions in machine ensemble theory and applications”, in M. Verleysen (ed.), Proc. 20th European Symp. on Artificial Neural Networks, Computational Intelligence, and Machine Learning, pp. 227-232; Lovaine-la-Neuve (Belgium): CIACO; 2012;

se exponen – breve y desordenadamente – posibilidades básicas y aplicadas.



15. Bibliografía (3)

La idea del énfasis mixto se publicó en dos versiones:

- V. Gómez-Verdejo, M. Ortega-Moral, J. Arenas-García, A.R. Figueiras-Vidal: “Boosting by weighting critical and erroneous samples”; Neurocomputing, vol. 69, pp. 679-685; 2006;
- V. Gómez-Verdejo, J. Arenas-García, A.R. Figueiras-Vidal, “A dynamically adjusted mixed emphasis method for building boosting ensembles; IEEE Trans. Neural Networks, vol. 19, pp. 3-17, 2008;

y se ha remitido a congreso (IWANN) la de ponderación suavizada:

- A. Ahachad, A. Omari, A.R. Figueiras-Vidal, “Smoothed emphasis for boosting ensembles”.

Los combinadores GG-FW aparecieron en:

- A. Omari, A.R. Figueiras-Vidal, “Feature combiners with gate generated weights for classification”, IEEE Trans. Neural Networks and Learning Systems, vol. 24, pp. 158-163; Jan. 2013.



15. Bibliografía (4)

El “boosting” con puertas, en:

- E. Mayhua-López, V. Gómez-Verdejo, A.R. Figueiras-Vidal, “Real AdaBoost with gate controlled fusion”, IEEE Trans. Neural Networks and Learning Systems, vol. 23, pp. 2003-2009; Dec. 2012.

El artículo dedicado a “boosting” con SVMs:

- E. Mayhua-López, V. Gómez-Verdejo, A.R. Figueiras-Vidal, “Boosting ensembles with subsampled LPSVM”,
está pendiente de publicación en la misma revista que las anteriores.

Una razonable revisión del aprendizaje profundo es:

- Y. Bengio, “Learning deep architectures for AI”, Foundations and Trends in Machine Learning, vol. 2, pp. 1-127, 2009;

y la propuesta de Hinton apareció en :

- G.E. Hinton, S. Osindero “A fast learning algorithm for deep belief nets”, Neural Computation, vol. 18, pp. 1527-1554, 2006.



15. Bibliografía (5)

Se están elaborando unos primeros artículos sobre agregaciones diversas, diversidad en aprendizaje profundo, y mejora/postagregación con GG-FW, y los trabajos en el resto de las líneas propias citadas progresa satisfactoriamente.

Finalmente: para quienes sientan curiosidad por la toma de decisiones por humanos y la inteligencia colectiva, los aconsejo que recurran a

- A.R. Figueiras-Vidal: *De máquinas y humanos. El arte de la toma de decisiones.* Discurso de Apertura del Año Académico 2013. Madrid: Real Academia de Ingeniería; 2013;

no porque se trate de un texto excepcional, sino por su seleccionada bibliografía de dichos temas.