










Ensamblado de Modelos Conceptos



Ensamblado: Concepto








En un concurso de la tele tenemos varios participantes. Se trata de acertar preguntas. Cuantas más aciertos, más dinero ganas. Veamos cómo lo hacen los participantes:

							
Pregunta 1	1	1	0	1	1	0	0
Pregunta 2	0	0	1	1	1	1	0
Pregunta 3	0	0	0	0	0	0	0
Pregunta 4	0	1	1	0	1	1	1
Pregunta 5	1	0	1	0	0	1	1
% ACIERTOS	40 %	40 %	60 %	40 %	60 %	60 %	40 %



Ensamblado: Concepto








En un concurso de la tele tenemos varios participantes. Se trata de acertar preguntas. Cuantas más aciertos, más dinero ganas. Veamos cómo lo hacen los participantes:

								VOTACIÓN
Pregunta 1	1	1	0	1	1	0	0	1
Pregunta 2	0	0	1	1	1	1	0	1
Pregunta 3	0	0	0	0	0	0	0	0
Pregunta 4	0	1	1	0	1	1	1	1
Pregunta 5	1	0	1	0	0	1	1	1
% ACIERTOS	40 %	40 %	60 %	40 %	60 %	60 %	40 %	80%



Ensamblado: Concepto

En un concurso de la tele tenemos varios participantes. Se trata de acertar preguntas. Cuantas más aciertos, más dinero ganas. Veamos cómo lo hacen los participantes:

								VOTACIÓN
Pregunta 1	1	1	0	1	1	0	0	1
Pregunta 2	0	0	1	1	1	1	0	1
Pregunta 3	0	0	0	0	0	0	0	0
Pregunta 4	0	1	1	0	1	1	1	1
Pregunta 5	1	0	1	0	0	1	1	1
% ACIERTOS	40 %	40 %	60 %	40 %	60 %	60 %	40 %	80%

¿Resultado? 7 concursantes trabajan mejor en equipo que de manera individual.
Este mismo comportamiento lo podemos extrapolar a los ensemblings



Ensamblado: Concepto

Los modelos ensamblados (ensemble models) combinan las decisiones de múltiples modelos para mejorar su precisión y estabilidad.

Se trata de modelos que se comportan muy bien y reducen bastante el variance (a costa del bias).

Este tipo de modelos son los que se suelen utilizar para ganar competiciones de Kaggle



Ensamblado: Concepto

Los modelos ensamblados (ensemble models) combinan las decisiones de múltiples modelos para mejorar su precisión y estabilidad.

Se trata de modelos que se comportan muy bien y reducen bastante el variance (a costa del bias).

Este tipo de modelos son los que se suelen utilizar para ganar competiciones de Kaggle

Tipos de ensembles más usados:

Bagging

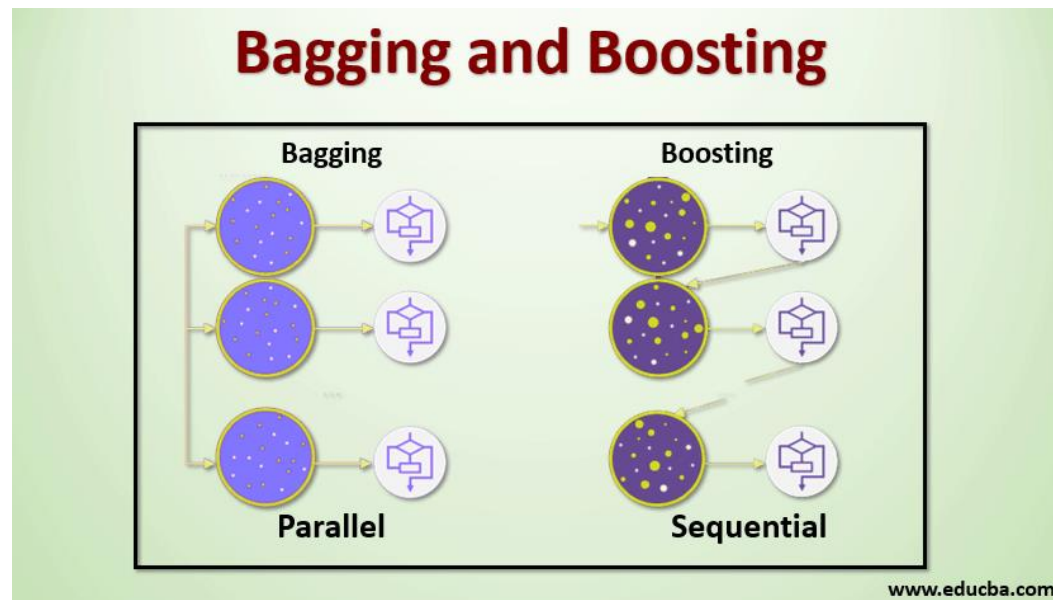
Random Forest

Boosting

AdaBoost

GradientBoost

XGBoost

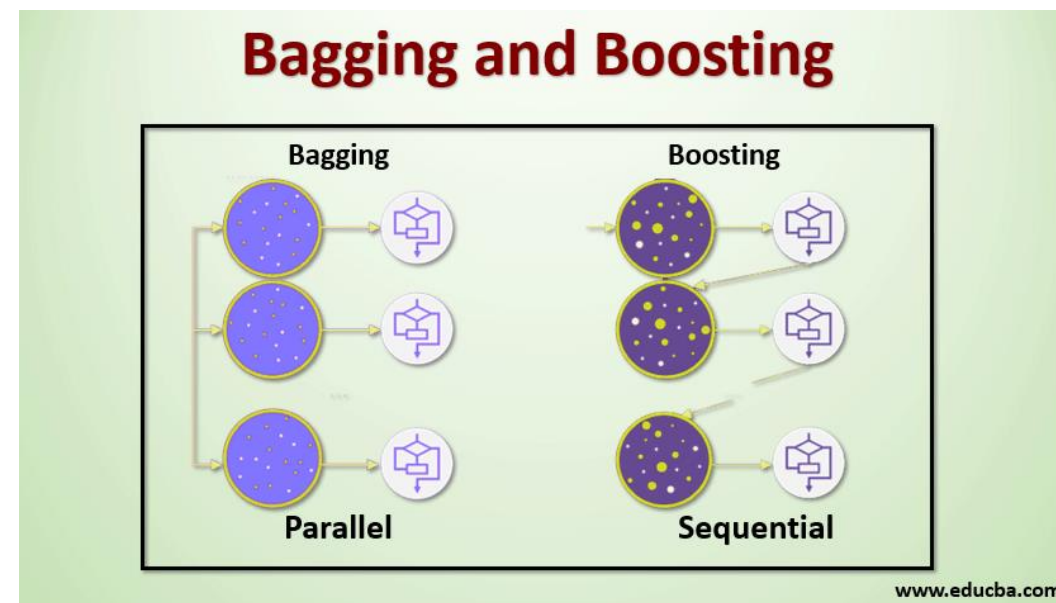


Ensamblado: Concepto

Los modelos pueden ser del mismo tipo o de tipos diferentes.

En algunos ensamblados los modelos se reparten el dataset de train (no ven todos los datos sino que se trocean y en algunos casos no ven todos las mismas features)

En otros los modelos no actúan en paralelo sino que actúan sobre las salidas de los modelos anteriores.



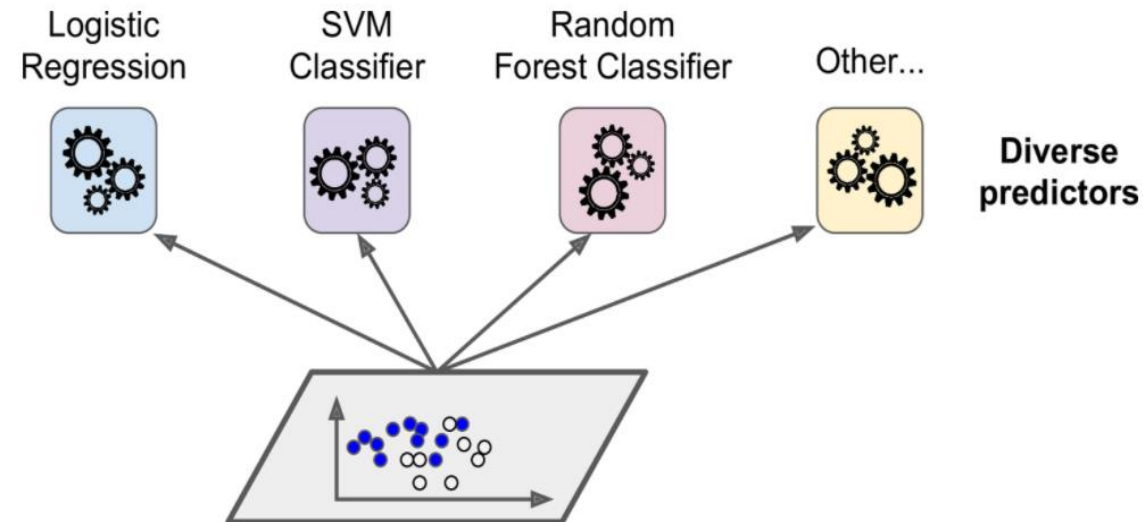
Ensamblado: Tipos

Distintos modelos

Se combinan modelos de varios tipos y luego se aplica un sistema de selección del conjunto de métricas.

Votación

Varios modelos trabajan en paralelo y se agregan sus outputs



Ensamblado: Tipos

Distintos modelos

Se combinan modelos de varios tipos y luego se aplica un sistema de selección del conjunto de métricas.



Votación

Varios modelos trabajan en paralelo y se agregan sus outputs

Mismos modelos

*Cada modelo del sistema de votación es del mismo tipo
(DecisionTree,
LogisticRegression...)*



Ensamblado: Tipos

Distintos modelos

Se combinan modelos de varios tipos y luego se aplica un sistema de selección del conjunto de métricas.

Votación

Varios modelos trabajan en paralelo y se agregan sus outputs

Mismos modelos

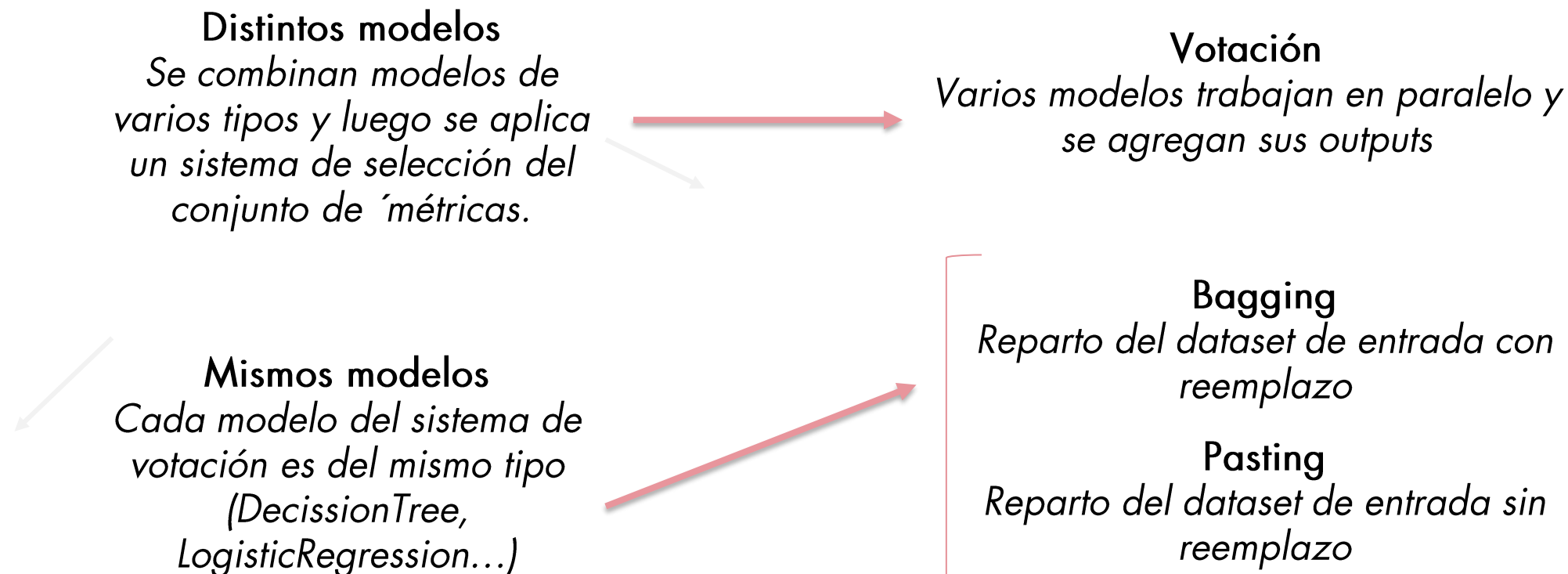
Cada modelo del sistema de votación es del mismo tipo (DecisionTree, LogisticRegression...)

Bagging

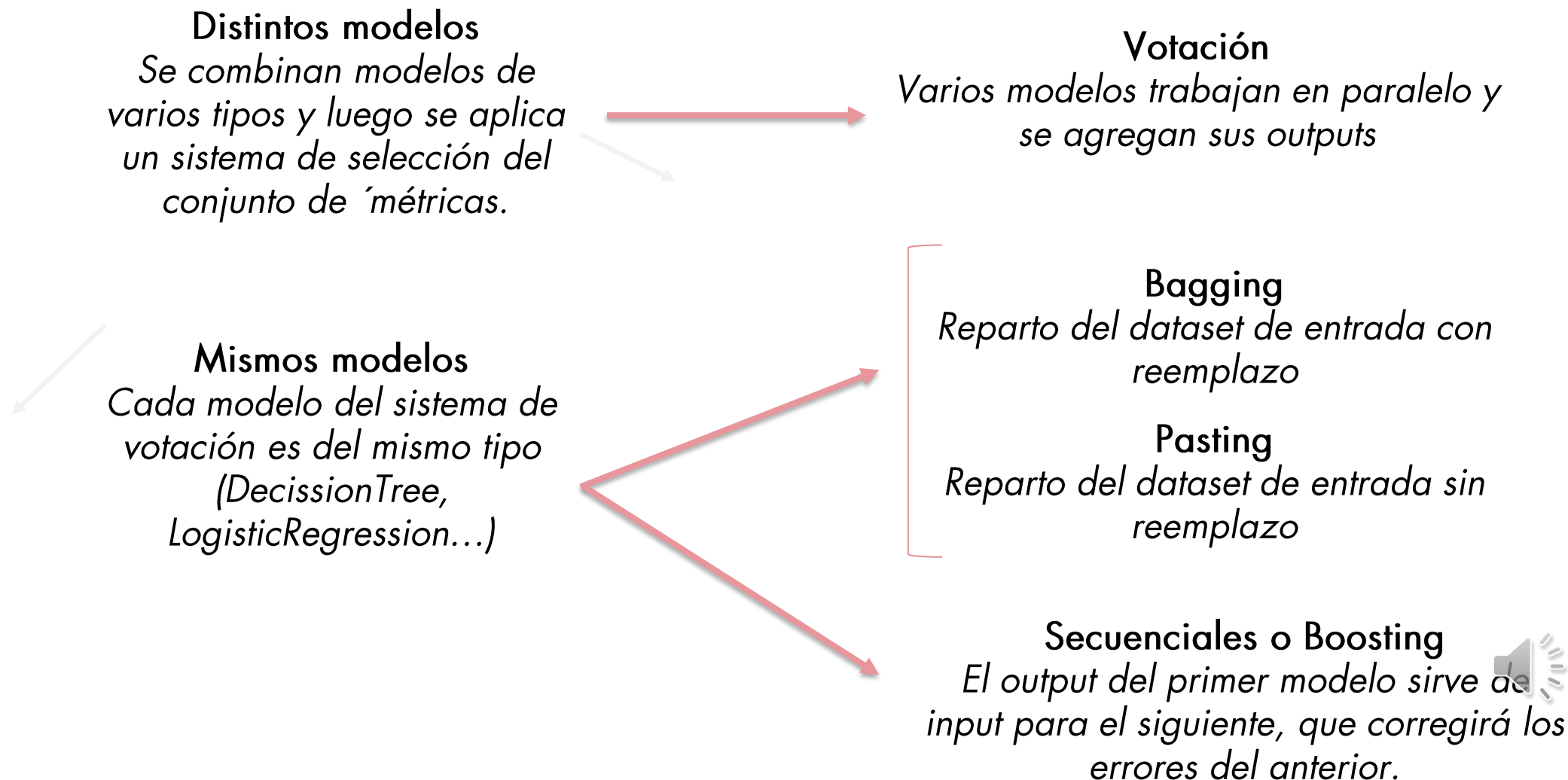
Reparto del dataset de entrada con reemplazo



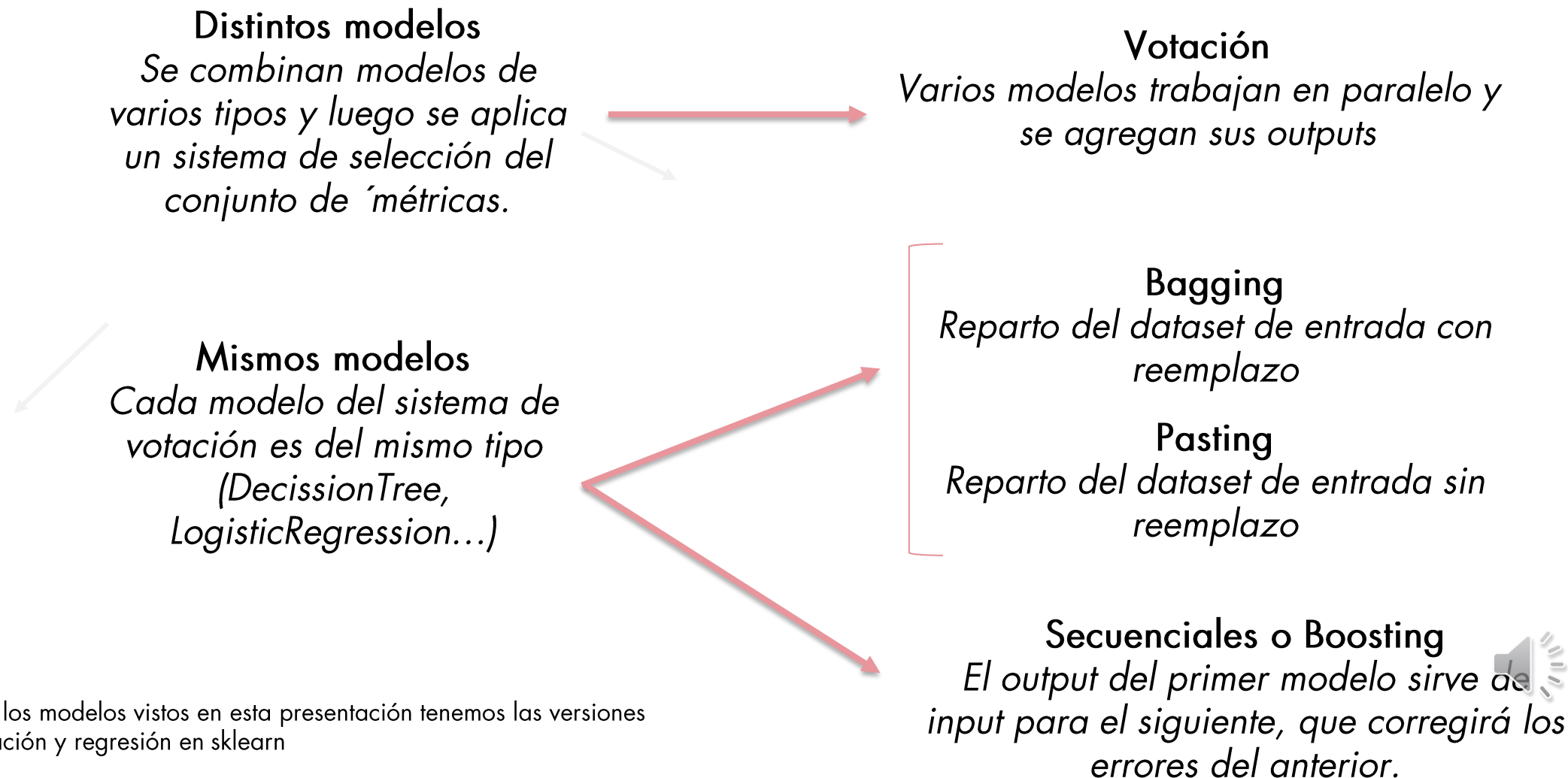
Ensamblado: Tipos



Ensamblado: Tipos



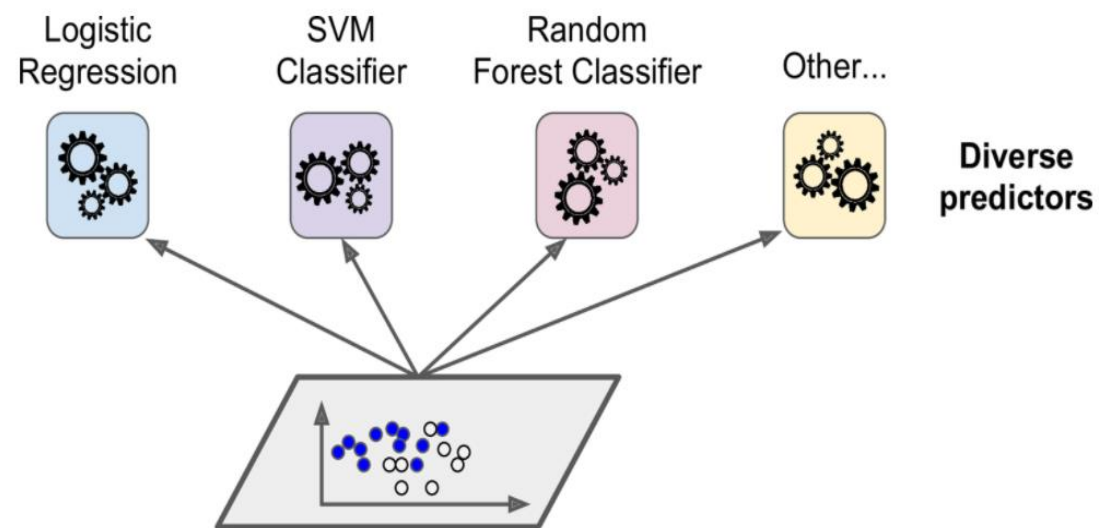
Ensamblado: Tipos



Para todos los modelos vistos en esta presentación tenemos las versiones de clasificación y regresión en sklearn

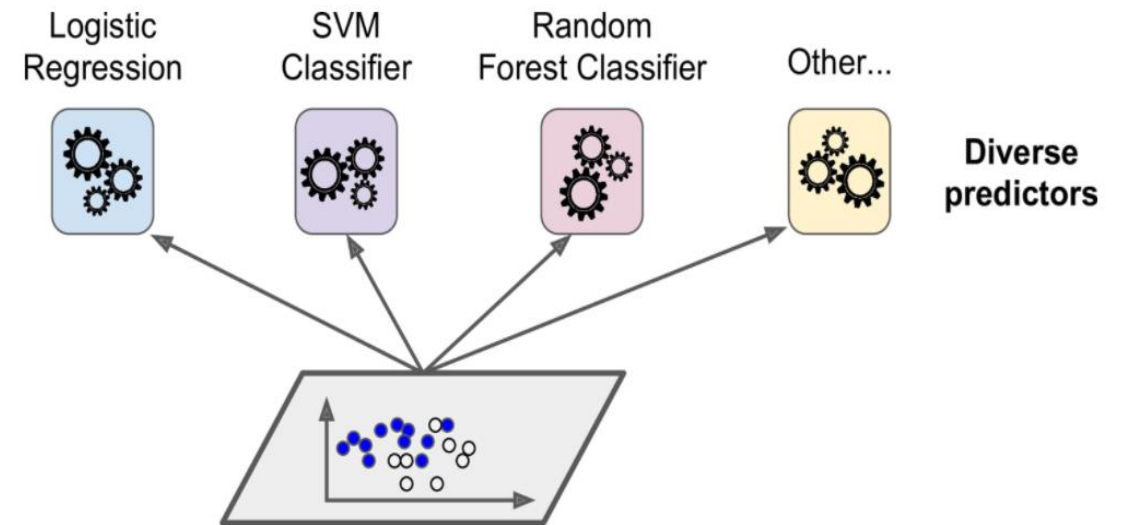
Ensamblado: Voting

- Veamos como funcionaría un ensamblado de modelos diversos con Voting



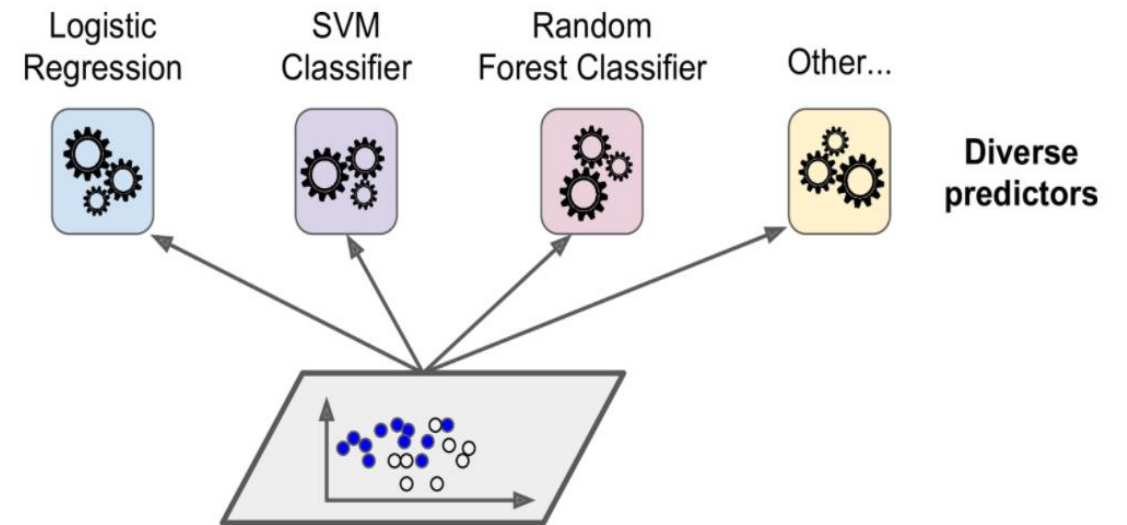
Ensamblado: Voting

- Veamos como funcionaría un ensamblado de modelos diversos con Voting
- 1. Escogemos nuestros modelos (que pueden ser diferentes o los mismos con diferentes hiperparámetros)



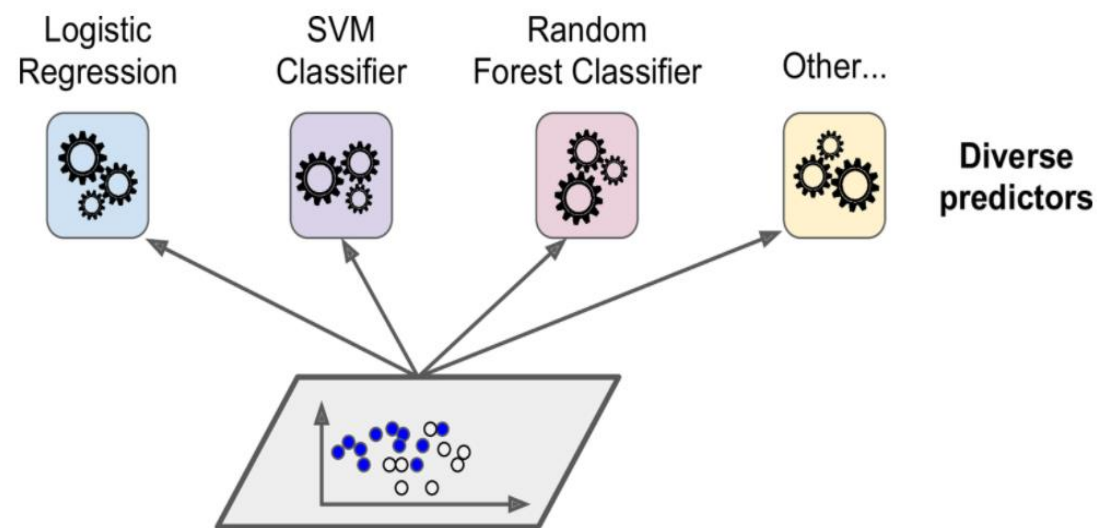
Ensamblado: Voting

- Veamos como funcionaría un ensamblado de modelos diversos con Voting
- 1. Escogemos nuestros modelos (que pueden ser diferentes o los mismos con diferentes hiperparámetros)
- 2. Entrenamos cada uno con el dataset de train



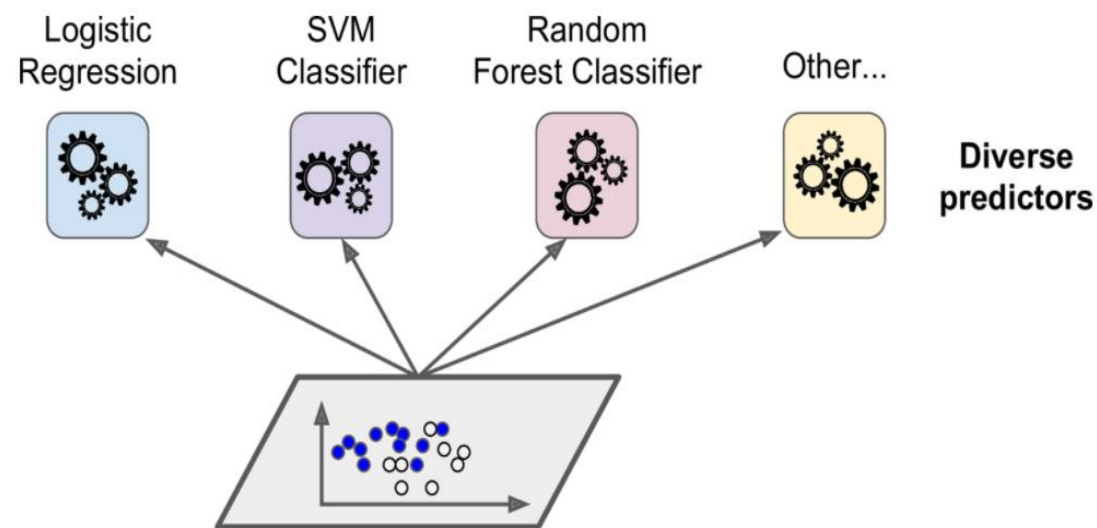
Ensamblado: Voting

- Veamos como funcionaría un ensamblado de modelos diversos con Voting
- 1. Escogemos nuestros modelos (que pueden ser diferentes o los mismos con diferentes hiperparámetros)
- 2. Entrenamos cada uno con el dataset de train
- 3. Cuando hay que predecir, todos los modelos predicen



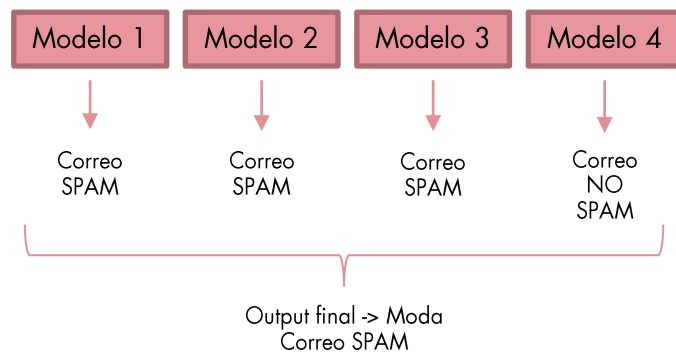
Ensamblado: Voting

- Veamos como funcionaría un ensamblado de modelos diversos con Voting
- 1. Escogemos nuestros modelos (que pueden ser diferentes o los mismos con diferentes hiperparámetros)
- 2. Entrenamos cada uno con el dataset de train
- 3. Cuando hay que predecir, todos los modelos predicen
- 4. Se usa un mecanismo de voting que sobre las predicciones de todos los modelos escoge la predicción final



Ensamblado: Voting

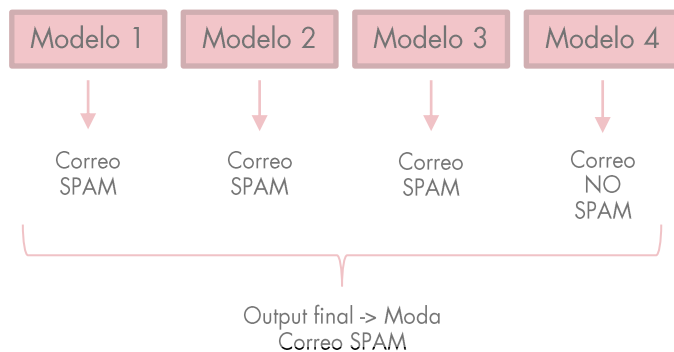
Clasificación – Hard voting
Se calcula la moda



Ensamblado: Voting

Clasificación – Hard voting

Se calcula la moda



Clasificación – Soft Voting

La clase que tenga la probabilidad más alta dentro de los outputs de todos los modelos

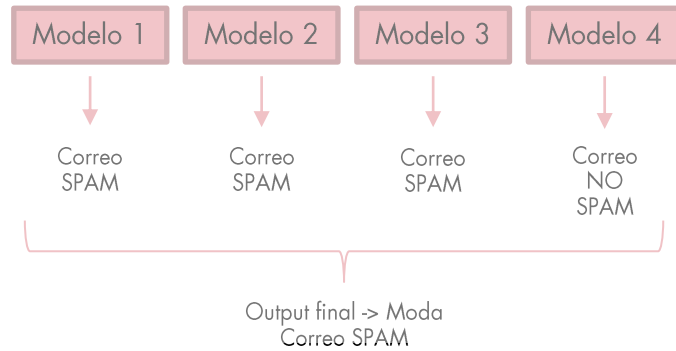
	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Media
SPAM	79%	71%	20%	85%	64%
NO SPAM	21%	29%	80%	15%	36%



Ensamblado: Voting

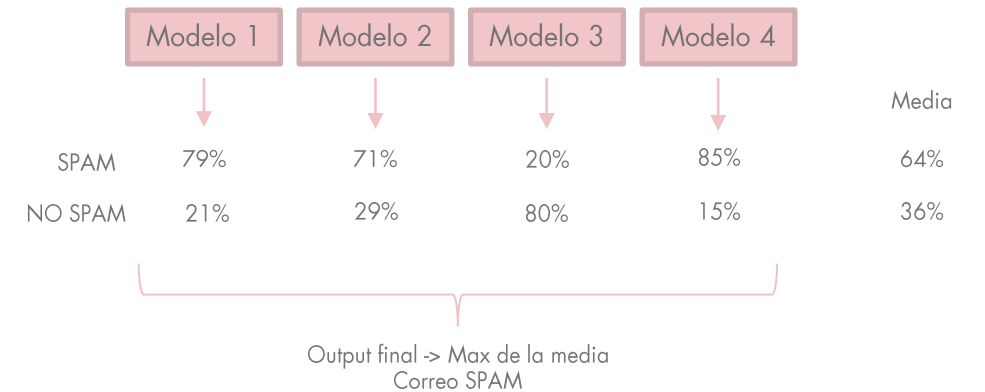
Clasificación – Hard voting

Se calcula la moda



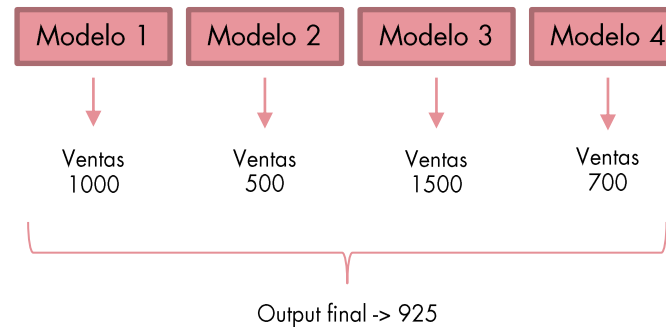
Clasificación – Soft Voting

La clase que tenga la probabilidad más alta dentro de los outputs de todos los modelos



Regresión

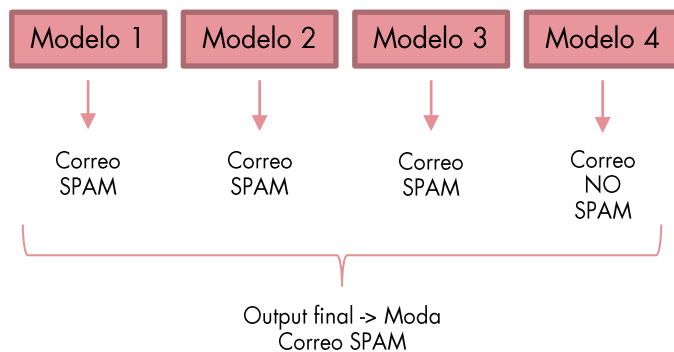
Se calcula la media de todos los outputs



Ensamblado: Voting

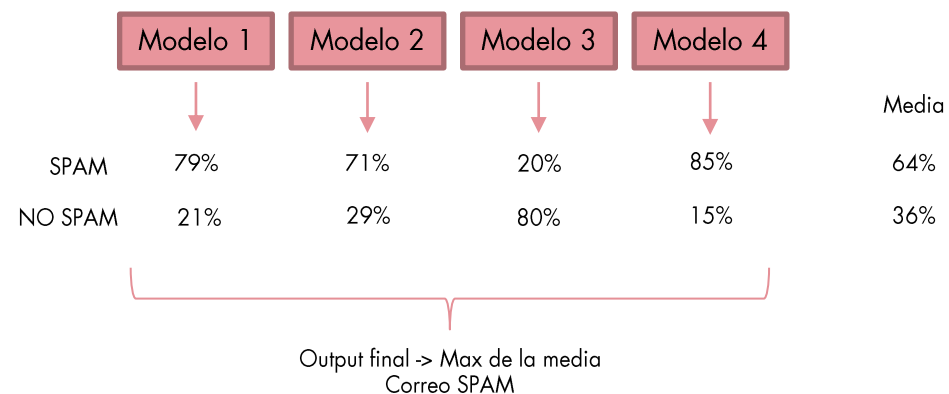
Clasificación – Hard voting

Se calcula la moda



Clasificación – Soft Voting

La clase que tenga la probabilidad más alta dentro de los outputs de todos los modelos



Regresión

Se calcula la media de todos los outputs

