

# Ensamble de clasificadores

Aprendizaje maquina

Dr. Rubén Acevedo  
ruben.acevedo@uner.edu.ar

Tecnicatura Universitaria  
Procesamiento y Exploración de Datos

# Ensamblados de clasificadores

## Introducción

Cuando se tiene que tomar decisiones médicas, financieras, otras; normalmente se toman en cuenta la opinión de varios expertos, tratando que esta sea la mejor posible.

En el contexto de aprendizaje maquinal, cada modelo inducido (clasificador) se puede ver como un *experto*, y en consecuencia se puede asumir que se produzcan predicciones más confiables si se combinan las predicciones de varios modelos

Desde 1990 aprox. hay disponibles técnicas para construir conjuntos de clasificadores cuyas decisiones son combinadas para clasificar nuevos ejemplos

# Ensamblas de clasificadores

## Introducción

Sistemas de múltiple clasificadores

Ensamble de clasificadores

Mezcla de expertos

Comité de clasificadores

# Ensamblajes de clasificadores

## Motivación

Razones estadísticas:

- Generalización de clasificadores individuales.
- Conjunto de clasificadores aumenta la generalización

Grandes volúmenes de datos:

- Entrenar un solo clasificador puede no ser práctico
- Dividir el conjunto de datos y entrenar clasificadores

Pocos datos:

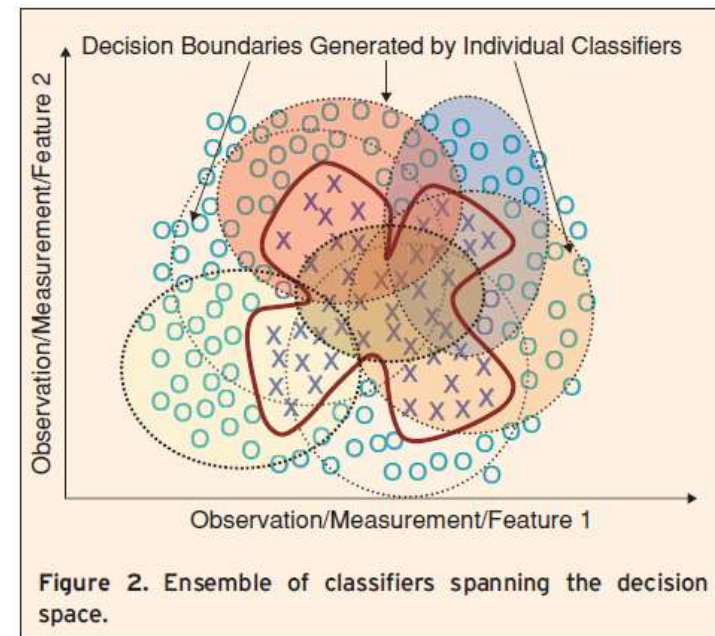
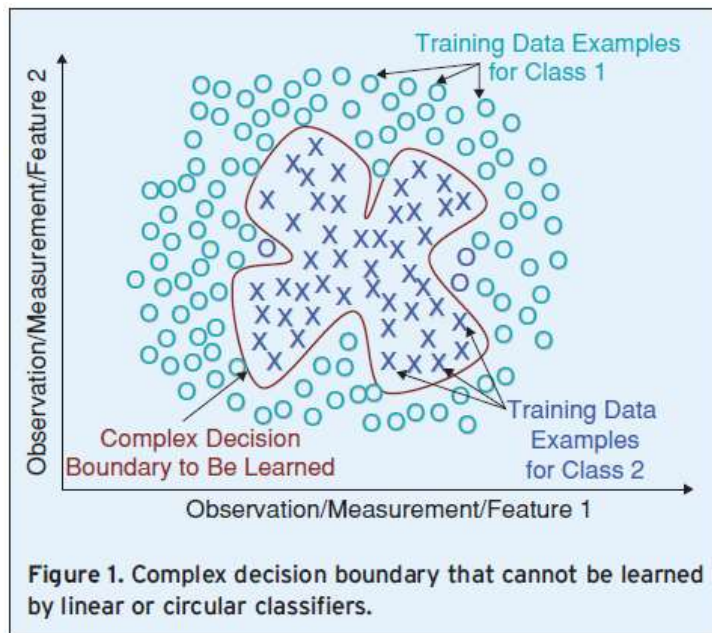
- Pobre representatividad del problema.
- Remuestreo de datos para entrenar clasificadores distintos.

# Ensamblados de clasificadores

## Motivación

Estrategia *divide y vencerás*:

- Regiones de decisión muy complicadas.
- Dividir el espacio en regiones mas sencillas de separar

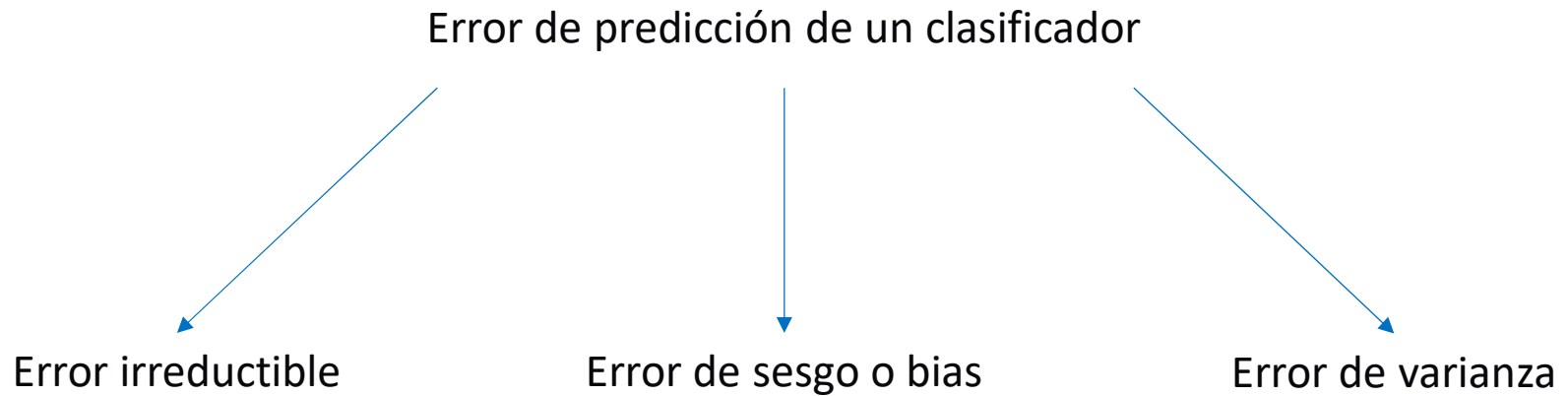


Adaptado de “Ensemble based systems in decision making”, IEEE Circuits and Systems Magazine, 2006.

# Ensamblas de clasificadores

Formalización

Enfoque estadístico



# Ensamblajes de clasificadores

Formalización

Enfoque estadístico

**Error irreducible:** no se puede reducir, independientemente de qué modelo (clasificador) se use.

También se le conoce como *ruido* y, por lo general, proviene por factores como variables desconocidas que influyen en el mapeo de las variables de entrada a la variable de salida, un conjunto de características incompleto o un problema mal enmarcado.

Es importante comprender que no importa cuán bueno hagamos nuestro modelo, nuestros datos tendrán cierta cantidad de ruido o un error irreducible que no se puede eliminar.

# Ensamblados de clasificadores

Formalización

Enfoque estadístico

**Error de sesgo:** Es la diferencia entre la predicción esperada del modelo y los valores verdaderos.

¿Cuál es el error inherente que obtiene un clasificador incluso con datos de entrenamiento infinitos? Esto se debe a que el clasificador está sesgado a un tipo particular de solución (por ejemplo, clasificador lineal).

Aunque al final el objetivo es siempre construir modelos que puedan predecir datos muy cercanos a los valores verdaderos, no siempre es tan fácil porque algunos algoritmos son simplemente demasiado rígidos para aprender señales complejas del conjunto de datos.

En otras palabras, *el sesgo es inherente al modelo/algoritmo*.



# Ensamblajes de clasificadores

Formalización

Enfoque estadístico

**Error de varianza:** captura cuánto cambia un clasificador si entrena con diferentes conjuntos de entrenamiento.

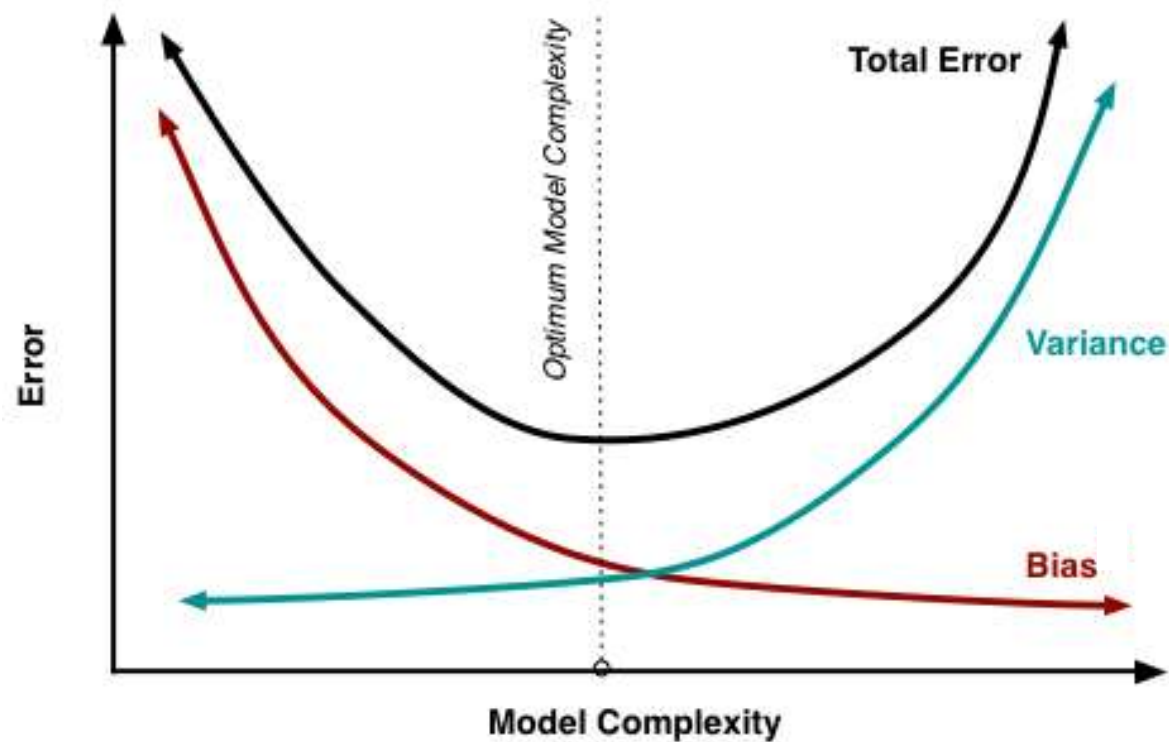
¿Qué tan especializado es un clasificador para un conjunto de entrenamiento en particular (*sobreajuste*)? Si tenemos el mejor modelo posible para nuestros datos de entrenamiento, ¿qué tan lejos estamos del clasificador promedio?

Idealmente no debería cambiar demasiado de un conjunto de datos de entrenamiento a otro, lo que significa que el algoritmo es bueno para elegir el mapeo subyacente oculto entre las variables de entrada y de salida.

# Ensamblados de clasificadores

Formalización

Enfoque estadístico

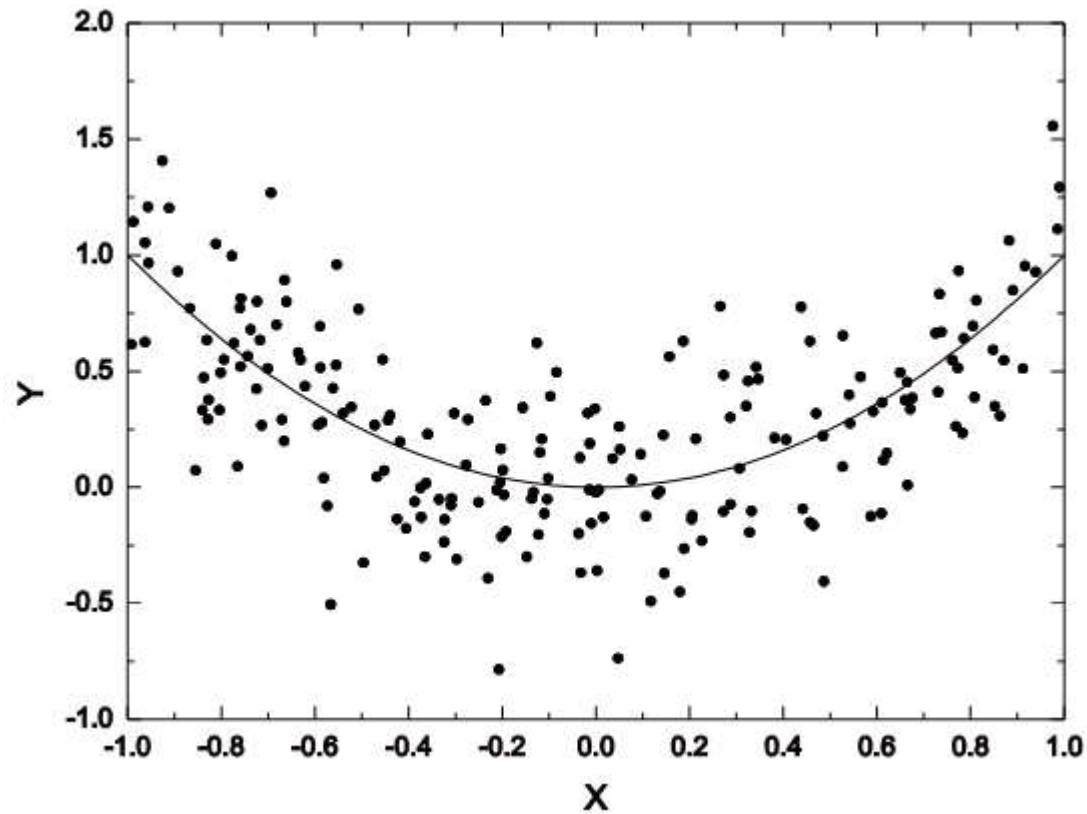


Los modelos más complejos se sobreajustan (error de varianza) mientras que los modelos más simples se subajustan (error de sesgo).

# Ensamblas de clasificadores

Formalización

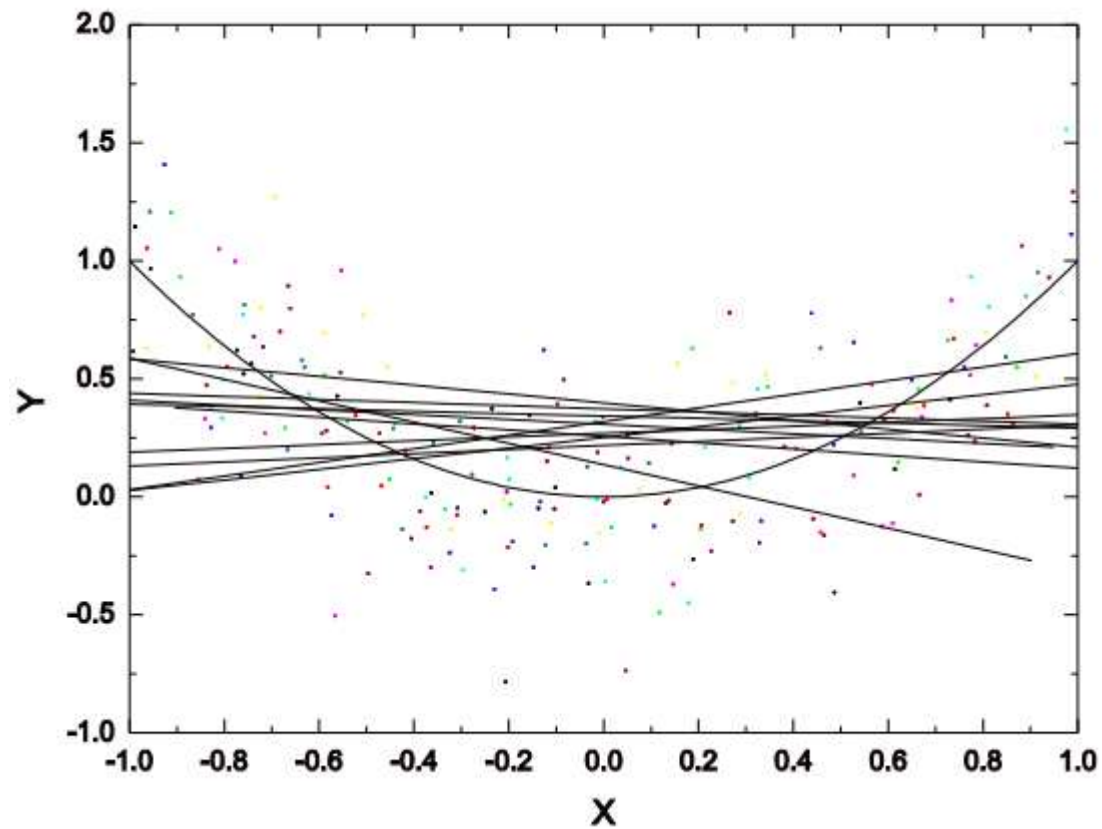
Enfoque estadístico



# Ensamblas de clasificadores

Formalización

Enfoque estadístico

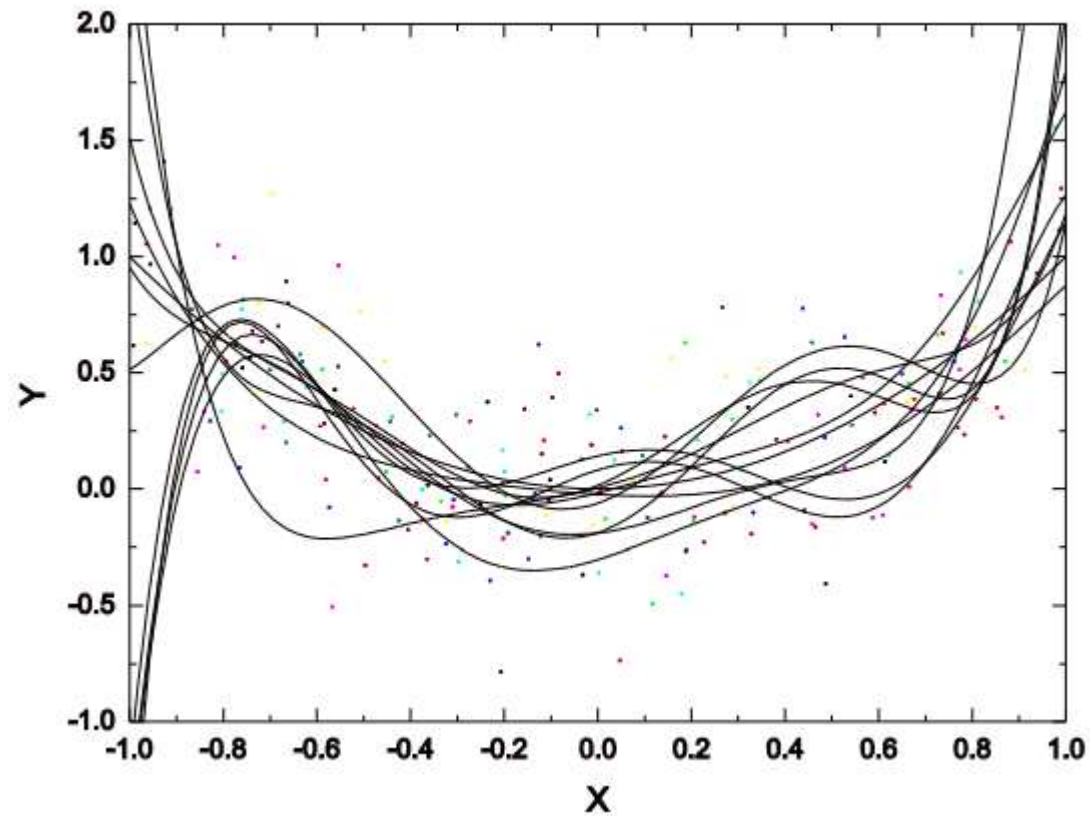


Error de sesgo

# Ensamblas de clasificadores

Formalización

Enfoque estadístico

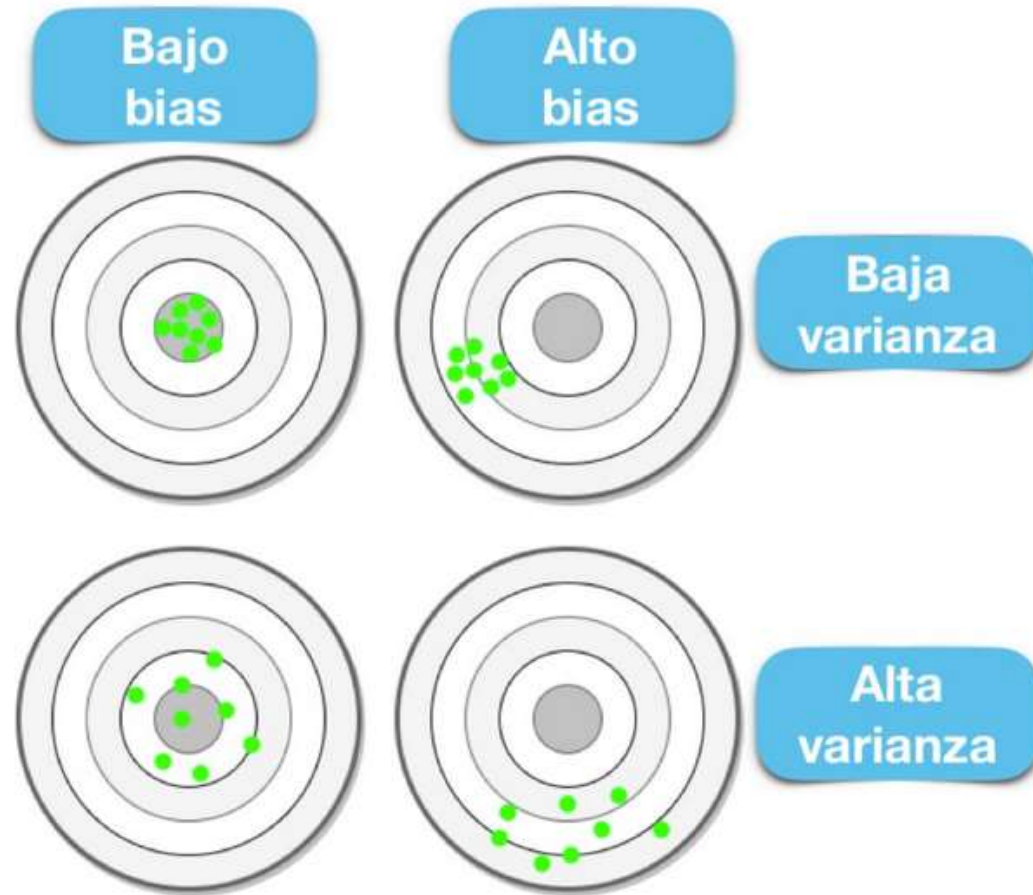


Error de varianza

# Ensamblajes de clasificadores

Formalización

Enfoque estadístico



# Ensamblados de clasificadores

Dilema...

Los predictores sin sesgo tienen alta varianza (y viceversa)

Hay dos formas de resolver el dilema:

Disminuir la varianza de los predictores sin sesgo (construir muchos predictores y promediarlos). Ejemplos: Bagging y Random Forest.

Reducir el sesgo de los predictores estables (construir una secuencia tal que la combinación tenga menos sesgo). Ejemplo: Boosting

# Ensamblados de clasificadores

## Diversidad

Si existiera un clasificador con generalización perfecta (para cualquier tipo de problema) no habría necesidad de recurrir a técnicas de ensamble. Sin embargo, efecto del ruido, los valores atípicos y la superposición distribuciones de datos; hacen que esto sea imposible.

El ensamble de clasificadores es una estrategia en la cual se combinan más de un clasificador, teniendo en cuenta la **diversidad** de sus respuestas, de forma tal que la combinación mejora el rendimiento de un solo clasificador.



# Ensamblajes de clasificadores

## Diversidad

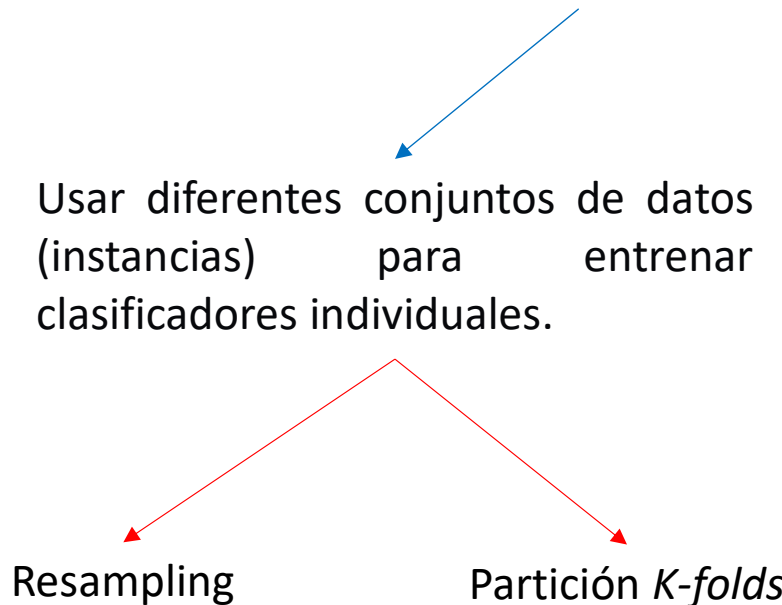
Esto requiere, que los clasificadores individualmente cometan errores sobre diferentes instancias. Intuitivamente, si cada clasificador comete errores diferentes, entonces una combinación estratégica de estos clasificadores puede reducir el error total.

Es deseable que cada clasificador sea único como sea posible, particularmente con respecto a las instancias clasificadas erróneamente. Por lo tanto, es necesario que las fronteras de decisión de cada clasificador sean diferentes de los de los demás, entonces dice que tal conjunto de clasificadores es **diverso**.

# Ensamblas de clasificadores

## Diversidad

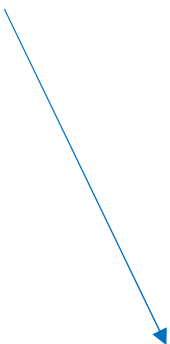
¿Como obtener la diversidad de clasificadores?



Usar diferentes conjuntos de datos  
(instancias) para entrenar  
clasificadores individuales.

Resampling

Partición *K-folds*



Usar diferentes parámetros de  
entrenamiento para diferentes  
clasificadores

# Ensamblajes de clasificadores

## Construcción

En la construcción de un ensamble se consideran dos componentes clave:

- 1) se requiere una estrategia de *combinación de clasificadores* para construir un ensamble tan diverso como sea posible.
- 2) es necesaria una estrategia para *combinar las salidas de los clasificadores* individuales que conforman el ensamble, de manera tal que las decisiones correctas sean reforzadas y las incorrectas se vean atenuadas o anuladas.

# Ensamblados de clasificadores

## Construcción

## Combinación clasificadores

- Bagging
- Variaciones de bagging:
  - Random forest
- Boosting
  - AdaBoost (Adaptive Boosting)
- Stacking

# Ensamblajes de clasificadores

Construcción

Combinación clasificadores

Bagging

En 1996, Breinmann propuso una estrategia denominada Bagging (*Bootstrap Aggregating*), que utiliza clasificadores del mismo tipo entrenados con distintos conjuntos de datos.

Funciona especialmente para algoritmos de aprendizaje inestables (cambian mucho sus estructuras al cambiar un poco los ejemplos), por ejemplo, los árboles de decisión y redes neuronales.

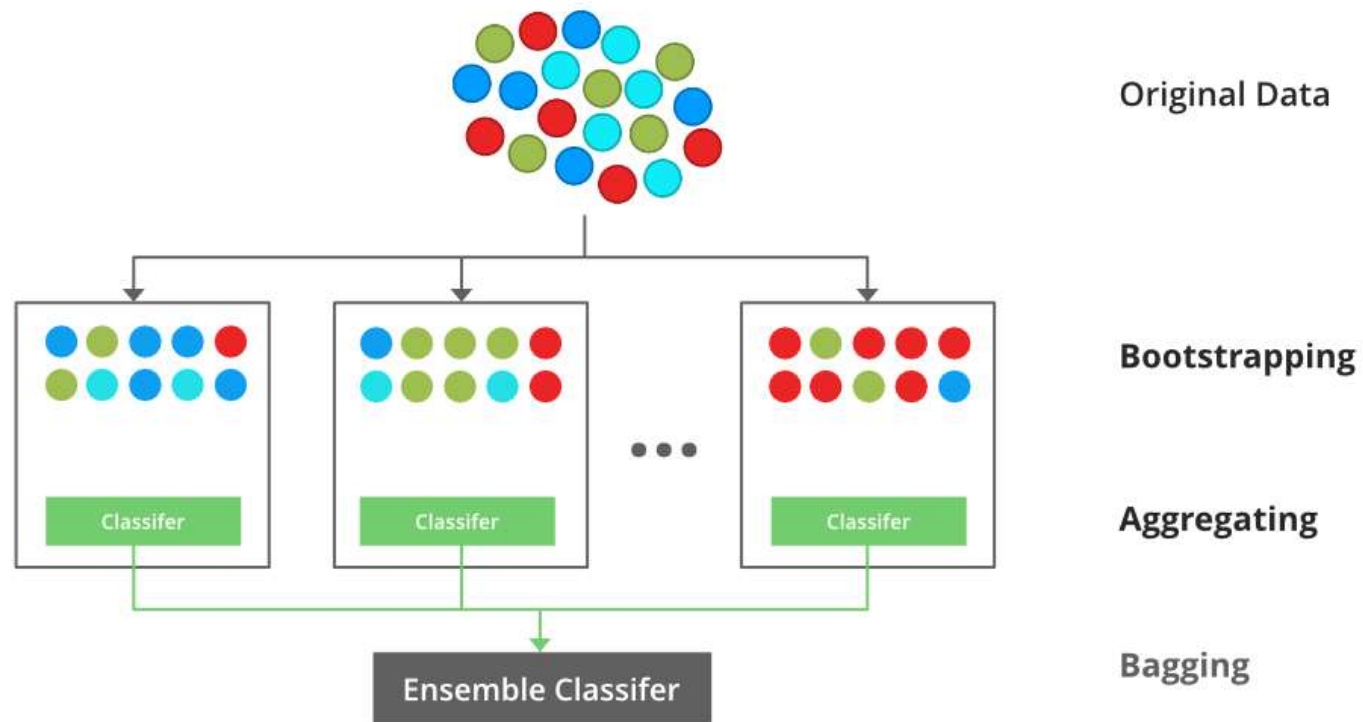
En general está orientado a problemas con escasez de datos.

# Ensamblados de clasificadores

Construcción

Combinación clasificadores

Bagging



# Ensamblados de clasificadores

Construcción

Combinación clasificadores

Bagging

## Implementación

Paso 1: Se crean múltiples subconjuntos de datos (instancias) a partir del conjunto de datos original, seleccionando datos con reemplazo.

Paso 2: Se crea un modelo base en cada uno de estos subconjuntos.

Paso 3: cada modelo se aprende en paralelo de cada conjunto de entrenamiento y de forma independiente entre sí.

Paso 4: Las predicciones finales se determinan combinando las predicciones de todos los modelos.

# Ensamblados de clasificadores

Construcción

Combinación clasificadores

Random forest

Random Forest (*bosque aleatorio*) es una variación de Bagging y está construido de árboles de decisión (AD).

Se construye con AD individuales cuyos parámetros de entrenamiento varían aleatoriamente.

Los AD crecen hasta separar todo, no hay podado ni criterio de parada.

El número de árboles no es importante mientras sean muchos (500, a 2000).



# Ensamblados de clasificadores

Construcción

Combinación clasificadores

Random forest

## Implementación

Paso 1: selección de  $k$  características de  $m$  totales (siendo  $k$  menor a  $m$ ) y creación de un árbol de decisión con esas  $k$  características.

Paso 2: creación de  $n$  AD variando siempre la cantidad de características y también podría variar la cantidad de datos para el muestreo por bootstrap.

Paso 3: cada uno de los  $n$  AD realiza la clasificación y se guardan los resultados.

Paso 4: cálculo de los votos obtenidos para cada “clase” seleccionada y se considera a la más votada como la clasificación final.

# Ensamblajes de clasificadores

Construcción

Combinación clasificadores

Boosting

En 1990, Schapire demostró que un algoritmo que combine clasificadores que simplemente pueden hacerlo mejor que adivinar al azar (débiles), se puede convertir en un algoritmo que genera un clasificador que puede clasificar correctamente todas las instancias de una fracción arbitrariamente pequeña del total.

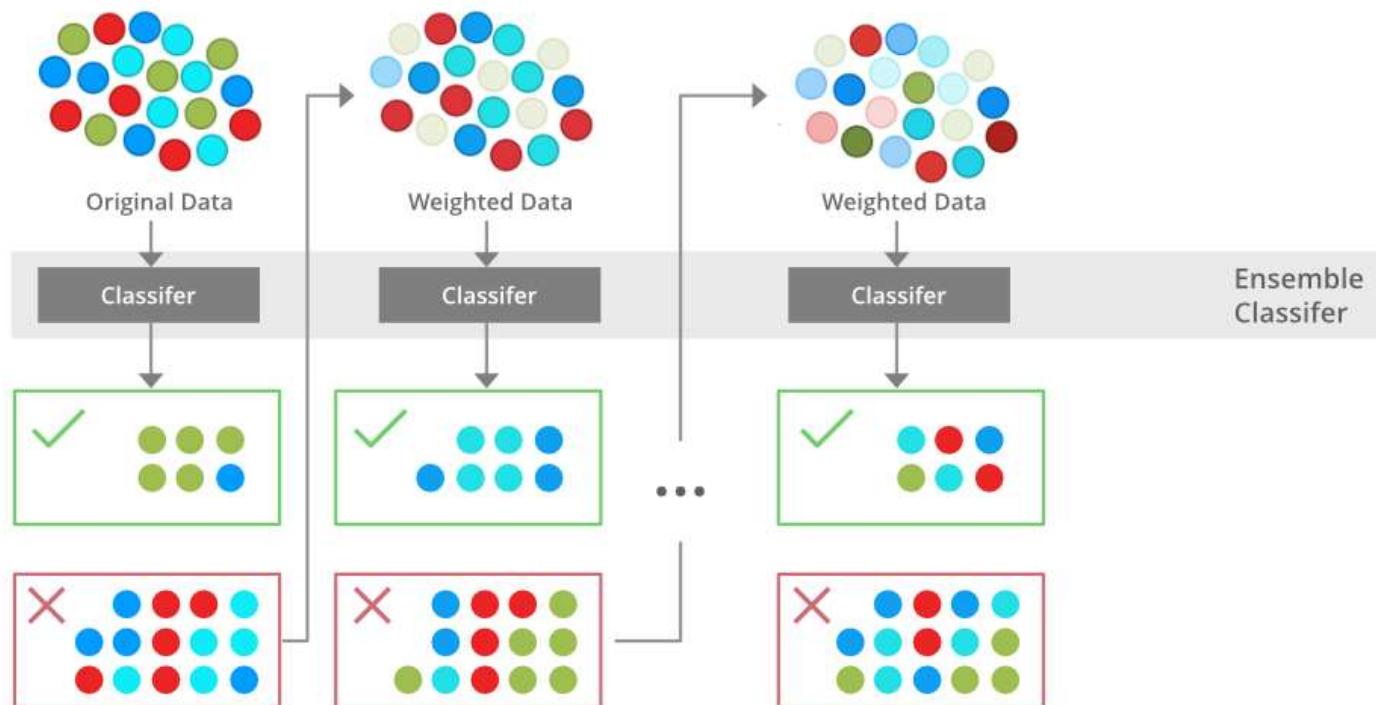
Demostró que el error de este esquema de tres clasificadores está acotado superiormente, y es menor que el error del mejor clasificador del conjunto.

# Ensamblados de clasificadores

Construcción

Combinación clasificadores

Boosting



# Ensamblajes de clasificadores

Construcción

Combinación clasificadores

Boosting

## Implementación (*AdaBoost*)

Paso 1: Inicializar el conjunto de datos y asignar el mismo peso a cada uno de los puntos de datos.

Paso 2: Proporcionar esto como entrada al modelo e identificar los datos clasificados incorrectamente.

Paso 3: Aumentar el peso de los datos mal clasificados.

Paso 4: *si* se obtuvieron los resultados requeridos

    Ir al Paso 5

*sino* Ir al Paso 2

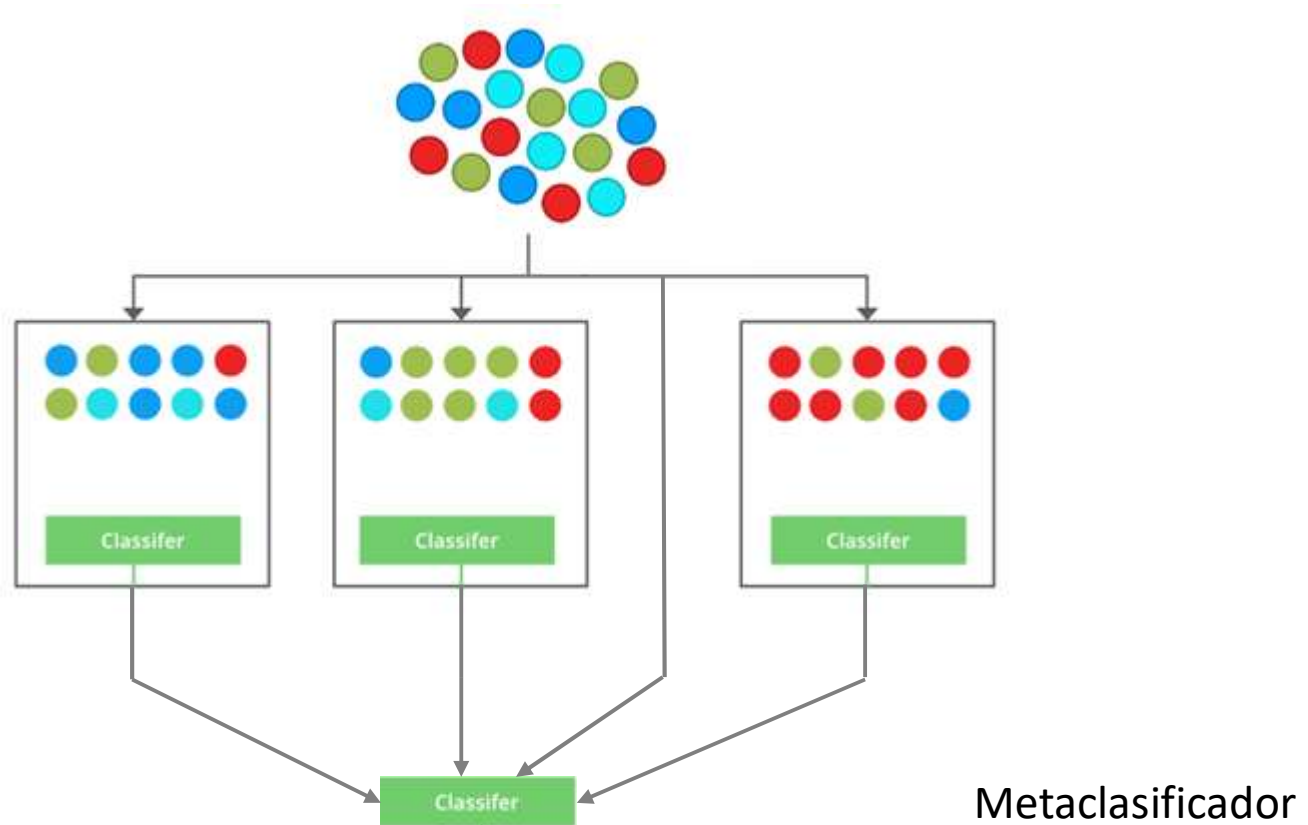
Paso 5: Fin

# Ensamblajes de clasificadores

Construcción

Combinación clasificadores

Stacking



# Ensamblajes de clasificadores

Construcción

Combinación de salidas

- Votación simple
- Votación ponderada
- Regla del mínimo/máximo/mediana
- Regla del producto
- Regla de la media

# Ensamblajes de clasificadores

Construcción

Combinación de salidas

Votación simple

- Votación unánime
- Votación por mayoría simple (mitad + 1)
- Votación por voto mayoritario

Votación pesada

- Votación ponderada (clasificadores mas calificados)

# Ensamblajes de clasificadores

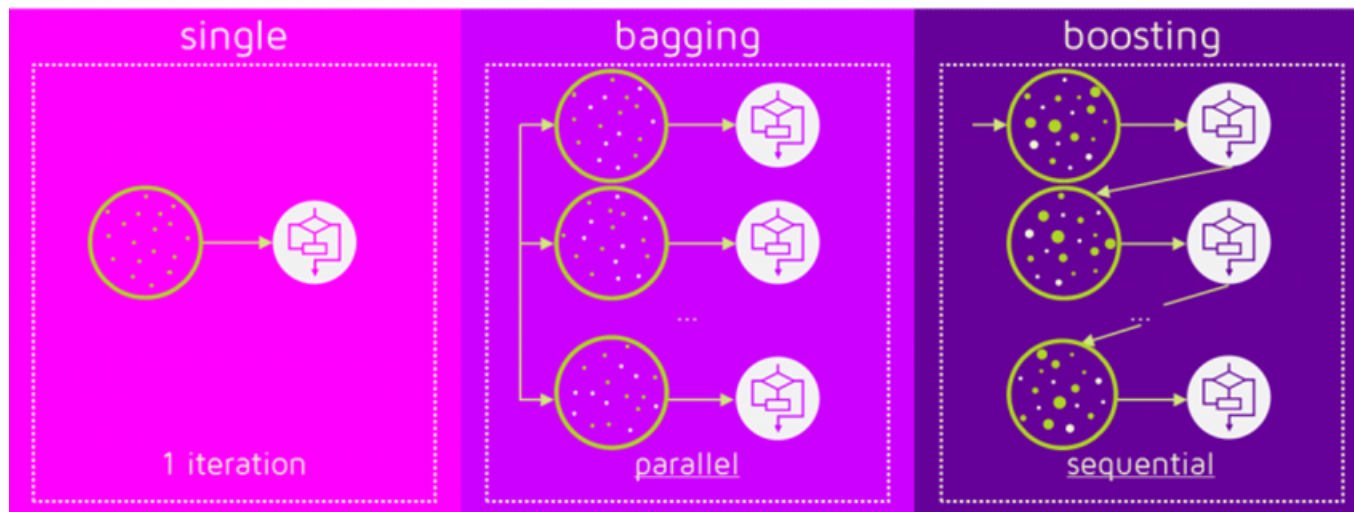
## Comparación

Bagging → clasificadores en paralelo

Bagging → resample

Boosting → clasificadores secuenciales

Boosting → resample datos pesados





# Ensamblajes de clasificadores

## Comparación

Bagging	Boosting
La forma más sencilla de combinar predicciones que pertenecen al mismo tipo.	Una forma de combinar predicciones que pertenecen a los diferentes tipos.
El objetivo es disminuir la varianza, no el sesgo.	El objetivo es disminuir el sesgo, no la varianza.
Cada modelo (clasificador) tienen el mismo peso.	Los modelos son pesados de acuerdo al desempeño.
Cada modelo se genera independientemente.	Los nuevos modelos son influenciados por el desempeño de los modelos anteriores.
Se extraen aleatoriamente diferentes subconjuntos de datos de entrenamiento con reemplazo del conjunto de datos de entrenamiento completo.	Cada nuevo subconjunto contiene los elementos que fueron mal clasificados por modelos anteriores.
Trata de resolver el problema de <i>overfitting</i> .	Trata de disminuir el sesgo

# Ensamblajes de clasificadores

## Consideraciones finales

¿Cuántos clasificadores utilizar?

- Decenas a centenas.

Bagging

- Es conveniente cuando los clasificadores son inestables (alta varianza).

Boosting

- Es conveniente cuando los clasificadores son estables (alto sesgo).

# Ensamblajes de clasificadores

Fin de la clase!