



Machine Learning

Unidad # 3 - Aprendizaje Supervisado Avanzado y Aprendizaje No Supervisado

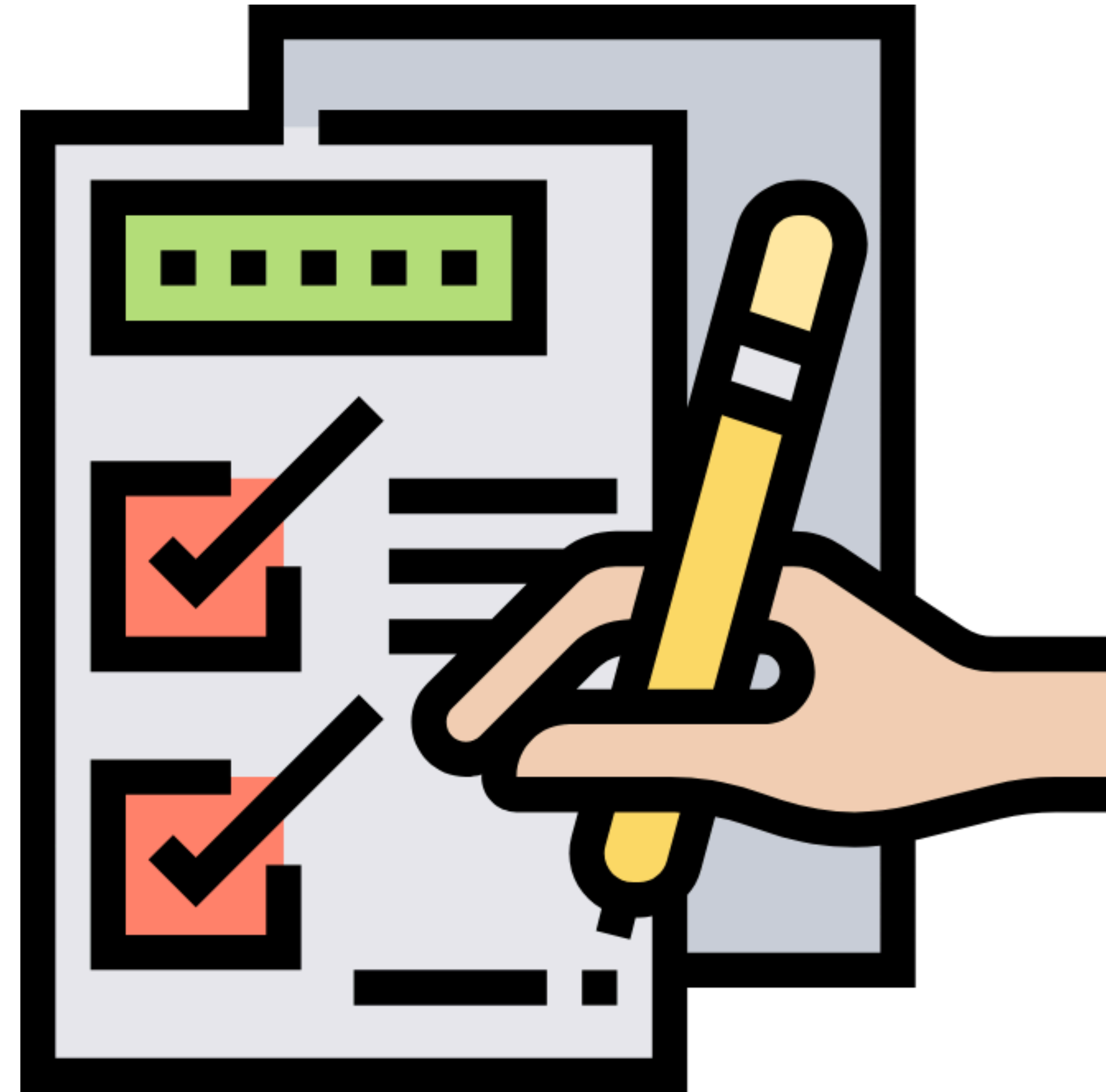
CC57 – 2019-1

Profesor
Andrés Melgar



Competencias a adquirir en la sesión

- Al finalizar la sesión el alumno comprenderá el funcionamiento del **aprendizaje inductivo**.
- Al finalizar la sesión el alumno combinará **modelos algorítmicos** usando conjuntos de datos.
- Al finalizar la sesión el alumno **entenderá** los métodos **ensembles**.
- Al finalizar la sesión el alumno **aplicará** los métodos **ensembles** para obtener modelos algorítmicos.





Métricas de Evaluación

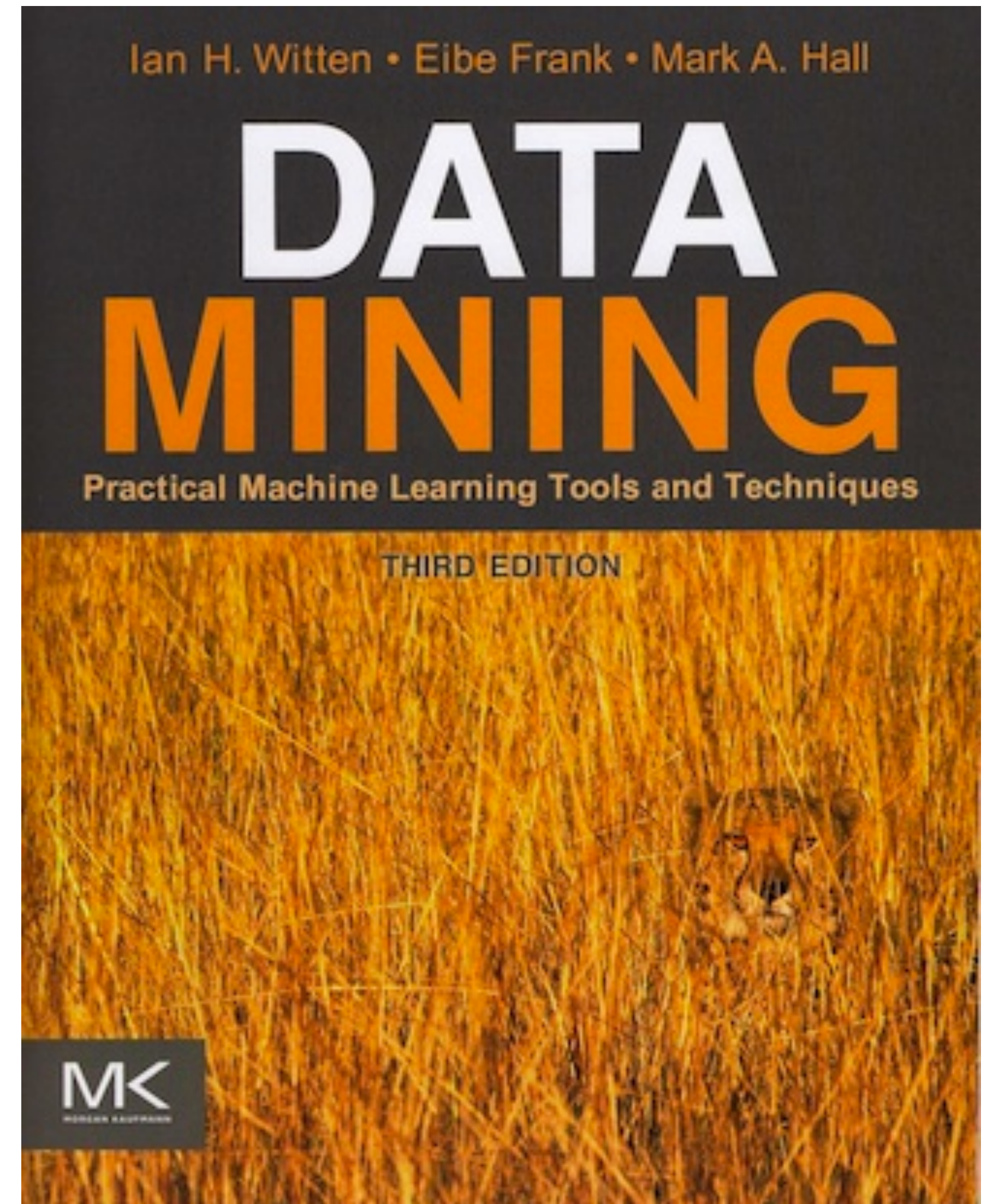
Texto guía

Witten, Ian H., Frank, Eibe, and Hall, Mark A.. 2011. *Data Mining : Practical Machine Learning Tools and Techniques with Java Implementations*. San Francisco: Elsevier Science & Technology.

CHAPTER

Ensemble Learning

8





Métodos Ensemble: Introducción

- Pasaremos a revisar técnicas que permiten **combinar** distintos modelos.
- A menudo es ventajoso tomar los datos de entrenamiento y
 - **Derivar** varios conjuntos de entrenamiento diferentes.
 - **Aprender** un modelo para cada uno.
 - **Combinarlos** para producir un conjunto de modelos.
- De hecho, las técnicas que permiten hacer esto son muy poderosas.
 - *Es posible transformar un esquema de aprendizaje débil en uno fuerte.*



Métodos Ensemble: Introducción

- La **pérdida de interpretabilidad** es un inconveniente cuando se aplican métodos ensemble, pero hay maneras para obtener descripciones estructuradas comprensibles basados en lo que aprenden estos métodos.
- Muchos de estos resultados son bastante contrario a la intuición, al menos a primera vista
 - ¿Cómo puede ser una buena idea usar modelos diferentes?
 - ¿Cómo se puede escoger el modelo que funciona mejor?
 - Sin duda, todo esto va en contra de la Navaja de Ockham, que aboga por la simplicidad
 - ¿Cómo se puede obtener un buen rendimiento combinando modelos diferentes?



Métodos Ensemble: Introducción

- Recuerde que en muchas ocasiones los **grupos** de humanos presentan decisiones más sabias que los individuos.
- Recordemos el **punto de vista de Epicuro** que, frente a las explicaciones alternativas, se deben mantener a todas.
- Imagine un **grupo de especialistas**. Cada especialista sobresale en un dominio. Pero ninguno tiene competencia en todos los ámbitos.
- Tratando de entender cómo funcionan estos métodos, los investigadores han expuesto todo tipo de conexiones y vínculos que han conducido a mejoras aún mayores.



Métodos Ensemble: Combinando múltiples modelos

- Cuando los sabios toman decisiones críticas, suelen tener en cuenta las **opiniones de varios expertos** en lugar de confiar en su propio juicio o la de un único asesor.
- Por ejemplo para tomar una decisión política importante.
 - Un dictador benigno consulta ampliamente, **no se sigue la opinión de un solo experto**.
 - En un entorno democrático, **la discusión de puntos de vista diferentes** puede producir un consenso; en caso contrario, se llama a elecciones.
 - En cualquier caso, se **combinan las diferentes opiniones** de los expertos.



Métodos Ensemble: Combinando múltiples modelos

- En la minería de datos, un modelo generado por el aprendizaje de máquina puede ser considerado como un experto.
- Un enfoque obvio para la toma de decisiones más fiable es la **combinación de la salida** de varios diferentes modelos.
- Varias técnicas de aprendizaje automático hacen esto mediante el **aprendizaje y combinación** de un conjunto de modelos.
- Entre estos podemos mencionar:
 - bagging
 - boosting
 - stacking.



Métodos Ensemble: Combinando múltiples modelos

- Bagging, boosting, y stacking se han desarrollado en el último par de décadas, y su rendimiento es a menudo sorprendentemente bueno.
- Investigadores del aprendizaje automático han tenido dificultades para entender por qué.
- Y durante esa lucha, nuevos métodos han surgido que a veces son incluso mejores.
- Mientras que los humanos rara vez se benefician del ruido, sacudir el modelo añadiendo variantes aleatorias de clasificadores puede mejorar el rendimiento.



Métodos Ensemble: Combinando múltiples modelos

- Un análisis más detallado revela que boosting, tal vez el más poderoso de los tres métodos, está estrechamente relacionado con la técnica estadística bien establecida de **modelos aditivos**, y este descubrimiento ha llevado a mejorar los procedimientos.
- Estos modelos combinados comparten la desventaja de ser bastante **difíciles de analizar**.
- Ellos comprenden decenas o incluso cientos de modelos individuales, y a pesar de que funcionan bien no es fácil de entender en términos intuitivos qué factores están contribuyendo a la mejora de las decisiones.



Métodos Ensemble: Bagging

- La combinación de las decisiones de los diferentes modelos significa la fusión de las distintas salidas en una única predicción.
- La forma más sencilla de hacer esto en el caso de la clasificación es tomar un voto (tal vez un voto ponderado).
- En el caso de la predicción numérica se calcula el promedio (tal vez una media ponderada).
- Tanto bagging como boosting adoptan este enfoque, pero derivan sus modelos individuales de diferentes maneras.
- En bagging los modelos reciben igual peso.
- En boosting se da más peso a los modelos más exitosos.



Métodos Ensemble: Bagging

- Introducción a bagging
 - Suponga que varios conjuntos de datos de entrenamiento del mismo tamaño son elegidos al azar de entre el dominio del problema
 - Imagine utilizar una técnica de aprendizaje máquina en particular para construir un árbol de decisión para cada conjunto de datos
- Se podría esperar que estos árboles sean prácticamente idénticos y hacer la misma predicción para cada instancia de una nueva prueba
- Pero, sorprendentemente, esta suposición no es generalmente cierta, sobre todo si los conjuntos de datos de entrenamiento son bastante pequeños.



Métodos Ensemble: Bagging

- La razón para ello es que la inducción de árbol de decisión es un proceso inestable.
- Ligeros cambios a los datos de entrenamiento puede fácilmente resultar en que un atributo diferente sea elegido en un nodo en particular.
- Esto podría generar otras ramificaciones para la estructura del subárbol debajo de ese nodo.
- Esto implica que hay casos de prueba para los que algunos de los árboles de decisión producen predicciones correctas y otras no.



Métodos Ensemble: Bagging

- Volviendo a la analogía de los expertos.
 - considere los expertos como los árboles de decisiones individuales.
 - Podemos combinar los árboles haciendo que voten en cada instancia de prueba.
 - Si una clase recibe más votos que cualquier otra, se toma como la correcta.
- En general, cuantos más mejor.
 - Las predicciones hechas por la votación se vuelven más fiable considerando los votos.
 - Las decisiones rara vez se deterioran si se descubren nuevos conjuntos de entrenamiento, los árboles están contruidos para ellos, y sus predicciones participan en la votación también.



Métodos Ensemble: Bagging. Bias–Variance Decomposition

- El efecto de la combinación de múltiples hipótesis puede ser visto a través de un modelo teórico conocido como la descomposición sesgo-varianza (Bias–Variance Decomposition).
- Supongamos que podríamos tener un número infinito de conjuntos de entrenamientos independientes de un mismo tamaño y los utilizamos para hacer un número infinito de clasificadores.
 - Una instancia de prueba es procesada por todos los clasificadores, y una sola respuesta se determina por mayoría de votos.
- En esta situación idealizada, todavía se producen errores porque ningún esquema de aprendizaje es perfecta
 - La tasa de error dependerá de lo bien que el método de aprendizaje de la máquina coincide con el problema en cuestión
 - También hay el efecto del ruido en los datos, que posiblemente no puede ser aprendido.



Métodos Ensemble: Bagging. Bias–Variance Decomposition

- Supongamos que la tasa de error esperada es evaluada promediando el error del clasificador combinado sobre un número infinito de ejemplos de prueba seleccionados independientemente.
- La tasa de error para un algoritmo de aprendizaje particular se llama su sesgo (bias) para el problema de aprendizaje y mide lo bien que el método de aprendizaje resuelve el problema.
- Mide el error "persistente" de un algoritmo de aprendizaje que no puede ser eliminado incluso mediante la adopción de un número infinito de conjuntos de entrenamiento.
- Por supuesto, no se puede calcular exactamente en situaciones prácticas; sólo puede ser se puede aproximar.



Métodos Ensemble: Bagging. Bias–Variance Decomposition

- Una segunda fuente de error en un modelo de aprendizaje, en una situación práctica, se deriva del conjunto de entrenamiento particular utilizado, que no es necesariamente finito y, por tanto, no representa a la población real de casos.
 - El valor esperado de este componente de error, sobre todos los posibles conjuntos de entrenamiento del tamaño dado y todos los posibles conjuntos de prueba, se llama la varianza del método de aprendizaje.
- El error total de este método de ensamble es compuesto por la suma del error del sesgo y el error en la varianza, lo que da el nombre al método descomposición sesgo-varianza.
- Estamos pasando por alto los detalles aquí.
 - La descomposición sesgo-varianza se introdujo en el contexto de la predicción numérica basado en error al cuadrado, donde hay una manera ampliamente aceptada de realizar la misma tarea
 - Sin embargo, la situación no es tan clara para la clasificación, y se han propuesto varias formas de realizar las descomposiciones



Métodos Ensemble: Bagging. Bias–Variance Decomposition

- Independientemente de la descomposición específica utilizada para analizar el error, la combinación de múltiples clasificadores disminuye el error esperado mediante la reducción de la varianza
- A mayor cantidad de clasificadores, mayor es la reducción de la varianza.
- Por supuesto, una dificultad surge al poner este sistema de votación en práctica
- Por lo general sólo hay un conjunto de entrenamiento, y la obtención de más datos es o bien imposible o costoso.



Métodos Ensemble: Bagging. Bias–Variance Decomposition

- Bagging intenta neutralizar la inestabilidad de los métodos de aprendizaje mediante la simulación del proceso descrito anteriormente utilizando un conjunto de entrenamiento dado.
- En lugar de un muestreo de los datos de entrenamiento independiente cada vez, los datos de entrenamiento originales se alteran mediante la supresión de algunos casos y se replican otros.
- Las instancias se muestrean al azar, con reemplazo, de la base de datos original para crear uno nuevo conjunto de datos del mismo tamaño. En este procedimiento de muestreo se replican inevitablemente algunos de los casos y se eliminan otros.



Métodos Ensemble: Bagging. Bias–Variance Decomposition

Model Generation

```
Let n be the number of instances in the training data.  
For each of t iterations:  
  Sample n instances with replacement from training data.  
  Apply the learning algorithm to the sample.  
  Store the resulting model.
```

Classification

```
For each of the t models:  
  Predict class of instance using model.  
Return class that has been predicted most often.
```




Métodos Ensemble: Boosting

- Hemos visto que bagging explota la inestabilidad inherente a los algoritmos de aprendizaje.
- Intuitivamente, la combinación de varios modelos sólo ayuda cuando estos modelos son significativamente diferentes entre sí y cada uno trata a un porcentaje razonable de los datos correctamente.
- Lo ideal es que los modelos se complementen entre sí, siendo cada uno un especialista en una parte del dominio en el que los otros modelos no funcionan muy bien.
- El método de boosting explota esta idea para combinar múltiples modelos mediante la búsqueda de forma explícita modelos que se complementan entre sí.



Métodos Ensemble: Boosting

- Similaridades con Bagging
 - Usa votos para clasificación y promedios para la predicción numérica para combinar la salida de modelos individuales.
 - Combina modelos del mismo tipo, por ejemplo árboles de decisión.
- Diferencias con Bagging
 - En bagging los diferentes modelos son contruidos de forma separada, en boosting cada modelos es influenciado por la performances de os modelos previos.
 - Boosting alienta a los nuevos modelos para convertirse en expertos para casos manejados incorrectamente por los anteriores mediante la asignación de un mayor peso a esas instancias.
 - Boosting no da la misma importancia a cada modelo, da un peso de acuerdo a su confianza.



Métodos Ensemble: Boosting.

AdaBoost.

- Hay muchas variantes de la idea de boosting. A continuación describiremos un método ampliamente utilizado llamado AdaBoost.M1 que está diseñado específicamente para la clasificación.
 - De forma análoga a bagging, se puede aplicar a cualquier algoritmo de clasificación.
- Para simplificar las cosas asumimos que el algoritmo de aprendizaje puede manejar casos ponderados, donde el peso de una instancia es un número positivo.
 - La presencia de instancia pesos cambia la forma en que se calcula el error de un clasificador.
 - Es la suma de los pesos de los casos mal clasificados dividido por el peso total de todos los casos, en lugar de la fracción de instancias que se clasificó erróneamente.



Métodos Ensemble: Boosting. AdaBoost.

- Por casos de ponderación, el algoritmo de aprendizaje puede ser obligado a concentrarse en un conjunto particular de casos, es decir, aquellos con alto peso.
- Tales casos se convierten en particularmente importante porque hay un mayor incentivo para clasificar correctamente.



Métodos Ensemble: Boosting.

AdaBoost.

- El algoritmo de boosting se inicia mediante la asignación de los mismos pesos a todas las instancias de entrenamiento.
- A continuación se llama al algoritmo de aprendizaje para formar un clasificador para estas instancias y asigna nuevos pesos a cada instancia de acuerdo a la salida del clasificador.
 - El peso de casos clasificados correctamente disminuye, y la de los clasificados erróneamente se incrementa.
 - Esto produce un conjunto de instancias "fáciles" con bajo peso y un conjunto de los "difíciles" con alto peso.
 - En las siguientes iteraciones se construyen clasificadores para los datos reponderados, se centra en la clasificación de los casos "difíciles".



Métodos Ensemble: Boosting. AdaBoost.

- Entonces los pesos de las instancias se aumentan o disminuyen de acuerdo con la salida de este nuevo clasificador
 - Como resultado, algunos casos difíciles pueden convertirse en las más difíciles.
 - Casos más fáciles, pueden convertirse incluso aún más fáciles.
 - Por otro lado, otras instancias difíciles puede ser volverse más fáciles
 - Algunas más fáciles pueden transformarse en más difíciles.
 - Todas las posibilidades pueden ocurrir en la práctica.



Métodos Ensemble: Boosting. AdaBoost.

- Después de cada iteración, los pesos reflejan con qué frecuencia los casos han sido clasificados erróneamente por los clasificadores producidos hasta ahora.
- Al mantener una medida de la “dificultad” en cada instancia, este procedimiento ofrece una manera elegante de generar una serie de expertos que se complementan entre sí.



Métodos Ensemble: Boosting.

AdaBoost.

Model Generation

```
Assign equal weight to each training instance.  
For each of t iterations:  
  Apply learning algorithm to weighted dataset and store resulting  
  model.  
  Compute error e of model on weighted dataset and store error.  
  If e equal to zero, or e greater or equal to 0.5:  
    Terminate model generation.  
  For each instance in dataset:  
    If instance classified correctly by model:  
      Multiply weight of instance by  $e / (1 - e)$ .  
  Normalize weight of all instances.
```

Classification

```
Assign weight of zero to all classes.  
For each of the t (or less) models:  
  Add  $-\log(e / (1 - e))$  to weight of class predicted by model.  
Return class with highest weight.
```




Competencias a adquirir en la sesión

- Al finalizar la sesión el alumno comprenderá el funcionamiento del **aprendizaje inductivo**.
- Al finalizar la sesión el alumno combinará **modelos algorítmicos** usando conjuntos de datos.
- Al finalizar la sesión el alumno **entenderá** los métodos **ensembles**.
- Al finalizar la sesión el alumno **aplicará** los métodos **ensembles** para obtener modelos algorítmicos.

