



Université de Batna 2 (Mostefa Ben Boulaid)
Faculté de Technologie
Département de Génie Industriel



Apprentissage Machine cours 1 (Introduction)

Pr Hassen BOUZGOU

Promotion doctorat 2020-2021

Intelligence: La notion d'intelligence peut être définie de plusieurs façons.

→ *la capacité de prendre la bonne décision.*

Naturelle

- S'adapter à de nouvelles situations pour survivre / se reproduire.
- Schémas comportementaux variés et adaptés
- Évolution + apprentissage individuel

Artificielle

- Nécessite l'acquisition de connaissances

→ **APPRENTISSAGE**

En Anglais:

Machine Learning: Faire apprendre des machines

C'est quoi Learning?

Apprendre des structures précédemment inconnue, à partir d'exemples

= CAPTURER LES VARIATIONS

exemple:

vous voulez faire un modèle pour mesurer la taille des enfants en fonction de leurs âge \Rightarrow *capturer les variations de la taille en fonction de l'âge*

Programmation traditionnelle



Apprentissage Automatique



- **Apprentissage supervisé**
 - Les données d'apprentissage incluent les sorties souhaitées
- **Apprentissage non-supervisé**
 - Les données d'apprentissage n'incluent pas les sorties souhaitées
- **Apprentissage semi-supervisé**
 - Les données d'apprentissage incluent quelques sorties souhaitées
- **Apprentissage par renforcement**
 - apprendre les actions à prendre, à partir d'expériences, de façon à optimiser une récompense quantitative au cours du temps



7210414959
0690159784
9665407401
3134727121
1742351244



c'est une branche de l'intelligence artificielle qui fait largement appel aux techniques d'apprentissage automatique et aux statistiques.

Reconnaissance d'objets à partir d'images:

- Chiffres, lettres, logos
- Biométrie: visages, empreintes, iris, ...
- Classes d'objets (voitures, arbres, bateaux, ...)
- Objets spécifiques (un visage, des yeux, ...)

- **Reconnaissance de visage**
- **Reconnaissance de caractère**
- **Détection de spam**
- **Diagnostic médical:** des symptômes aux maladies
- **Biométrie:** Reconnaissance / authentification utilisant des caractéristiques physiques et / ou comportementales: Visage, iris, signature, etc.

De manière générale, nous aimerions prédire une valeur t à partir d'une observation x





$$t = y(x, w)$$

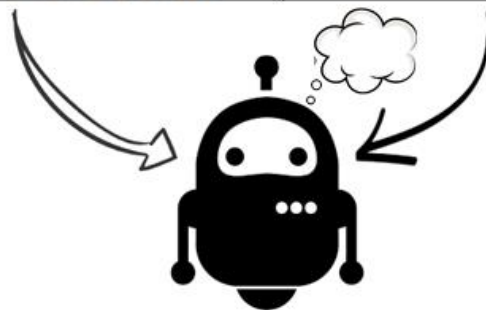
- Si y est continue : **régression**
- Si y est discrète : **classification**

Apprentissage des paramètres w à partir de données.

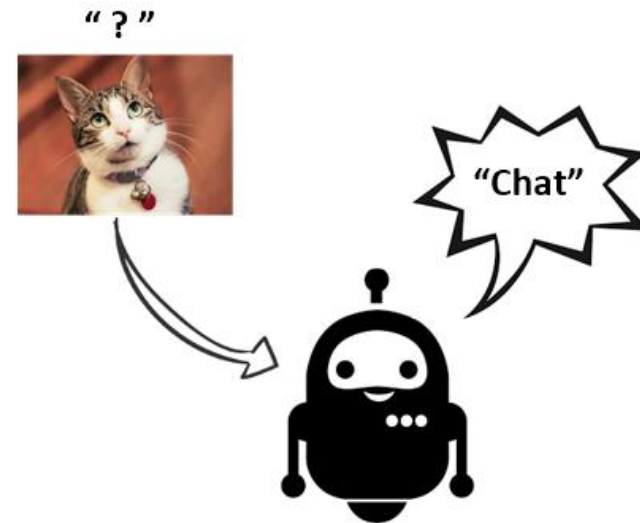
L'apprentissage supervisé (supervised learning en anglais) est une technique d'apprentissage automatique où l'on cherche à produire automatiquement des règles à partir d'une base de données d'apprentissage contenant des « exemples » (en général des cas déjà traités et validés).



x	y
	"Chien"
	"Chien"
	"Chat"
	"Chien"

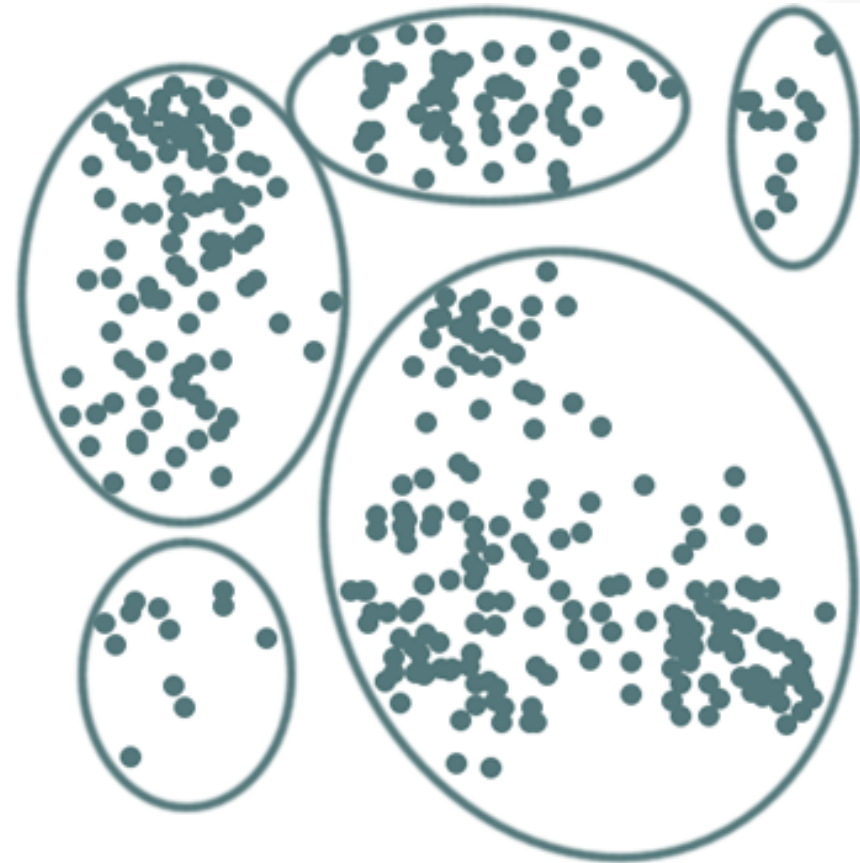


Apprentissage Supervisé

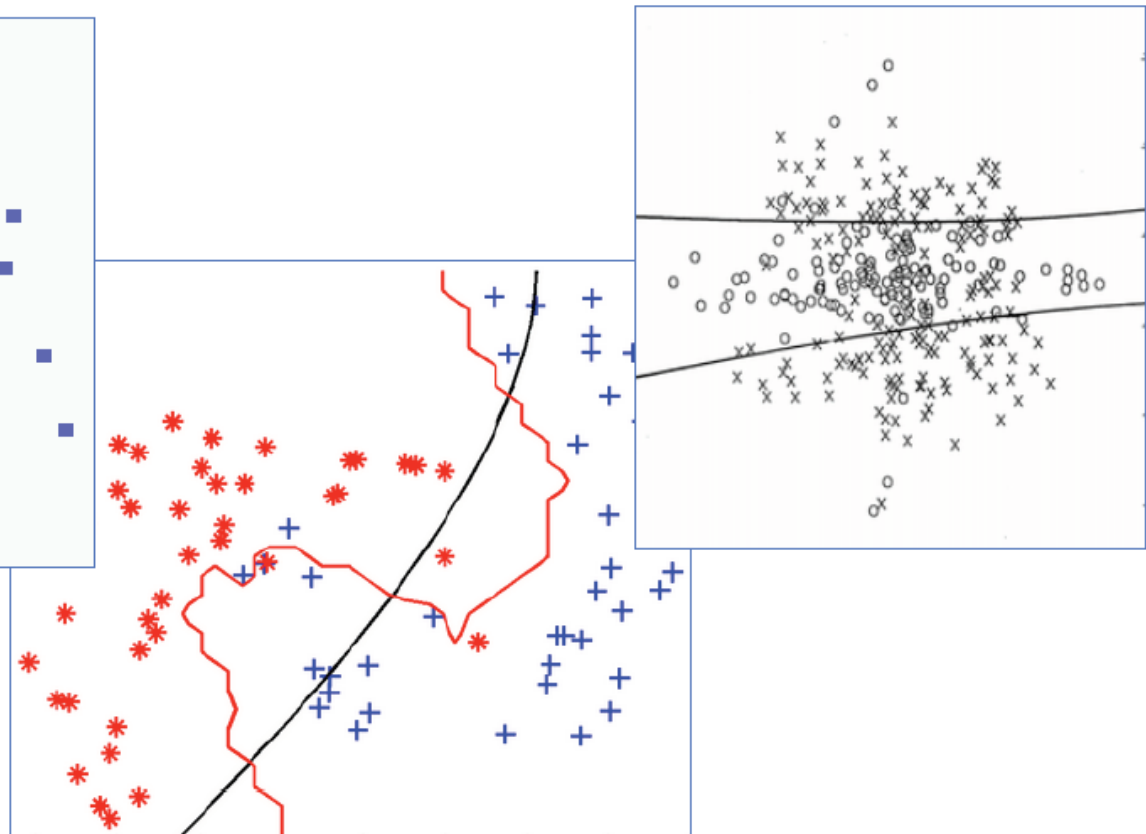
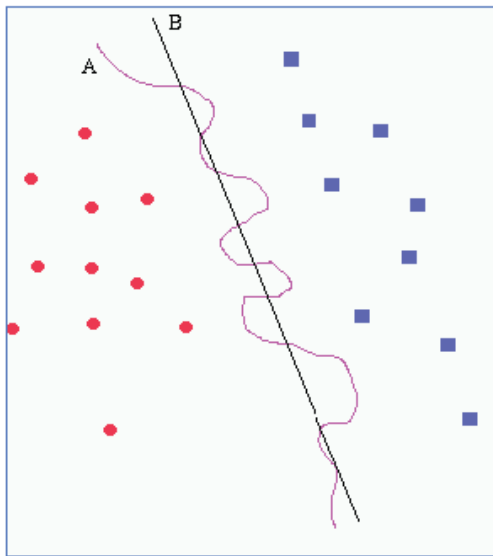


Utilisation finale

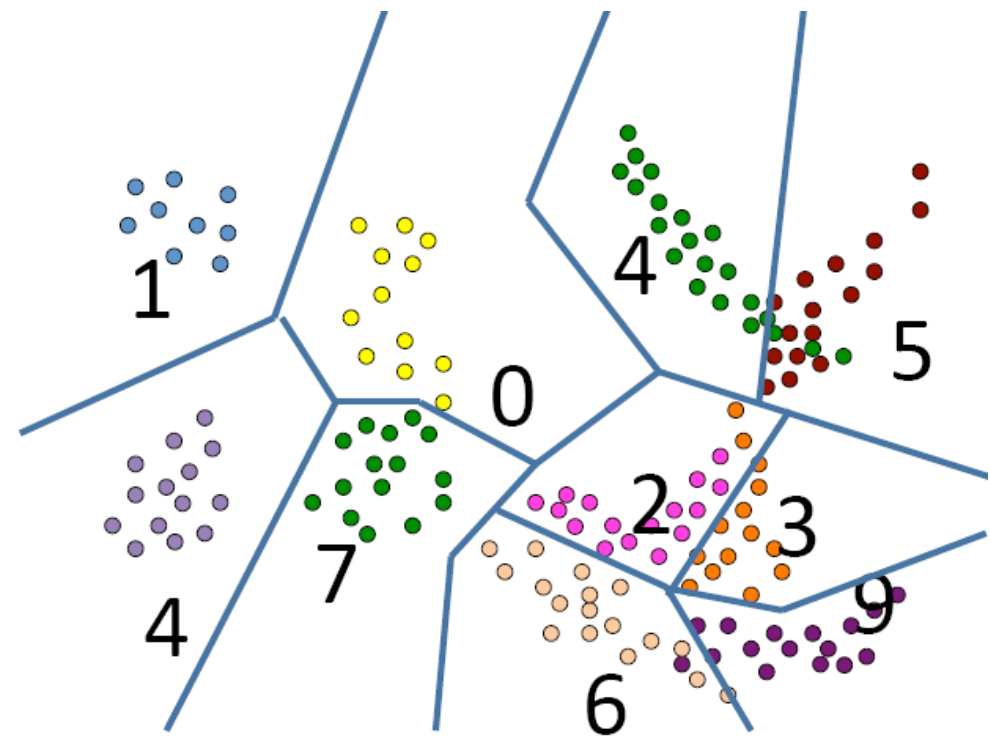
l'**apprentissage non supervisé** (parfois dénommé « **clustering** ») est une méthode d'apprentissage automatique. Il s'agit pour un logiciel de diviser un groupe hétérogène de données, en sous-groupes de manière que les données considérées comme les plus similaires soient associées au sein d'un groupe homogène et qu'au contraire les données considérées comme différentes se retrouvent dans d'autres groupes distincts ; l'objectif étant de permettre une extraction de connaissance organisée à partir de ces données.



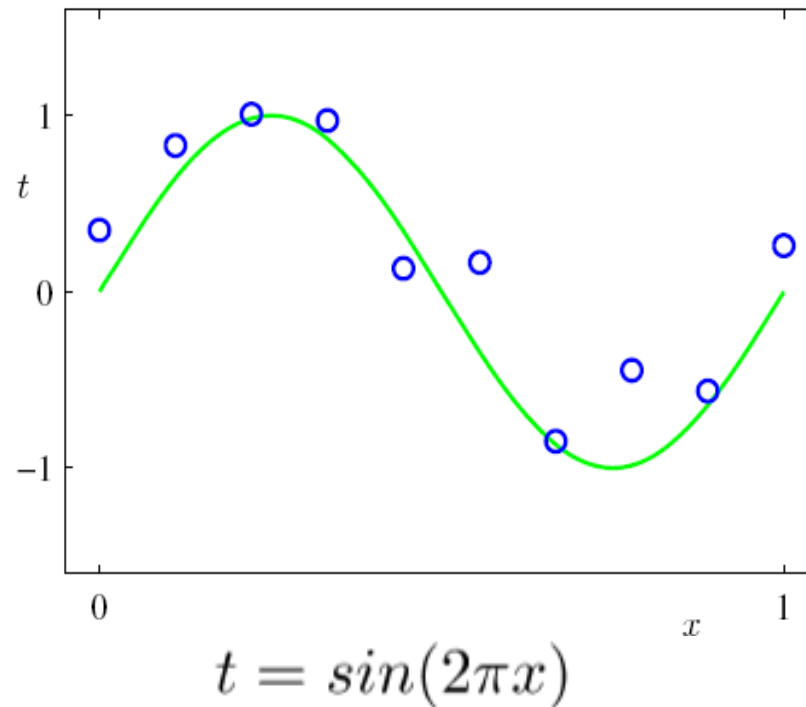
Apprendre à **classifier** des données revient à apprendre une **fonction de décision** : → la frontière entre les classes.



La complexité d'une fonction de décision dépend de la complexité du regroupement des étiquettes dans l'espace de caractéristiques (*feature space*)



- Données générées par une fonction $t = \sin(2\pi x)$
Objectif : supposant la fonction inconnue, prédire t à partir de x



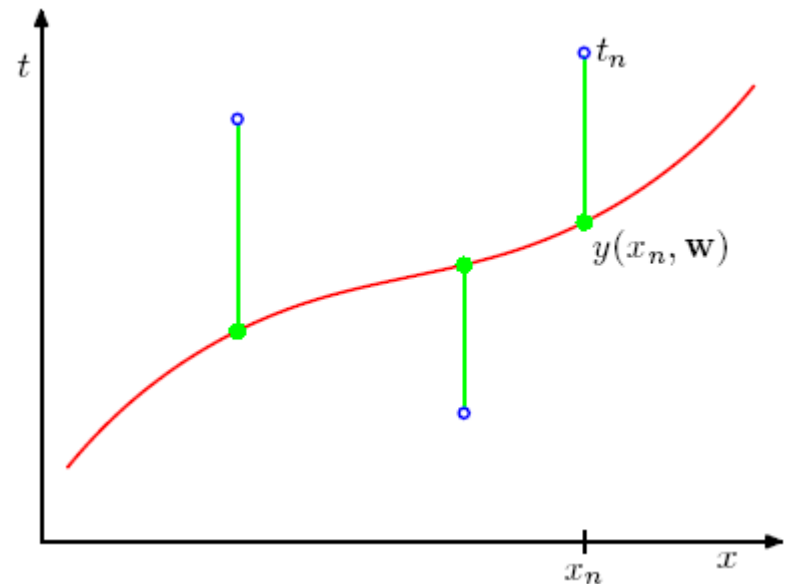
Fitting d'un polynôme d'ordre M

$$y(x, \mathbf{w}) = w_0 + w_1x + w_2x^2 + \dots + w_Mx^M = \sum_{j=0}^M w_jx^j$$

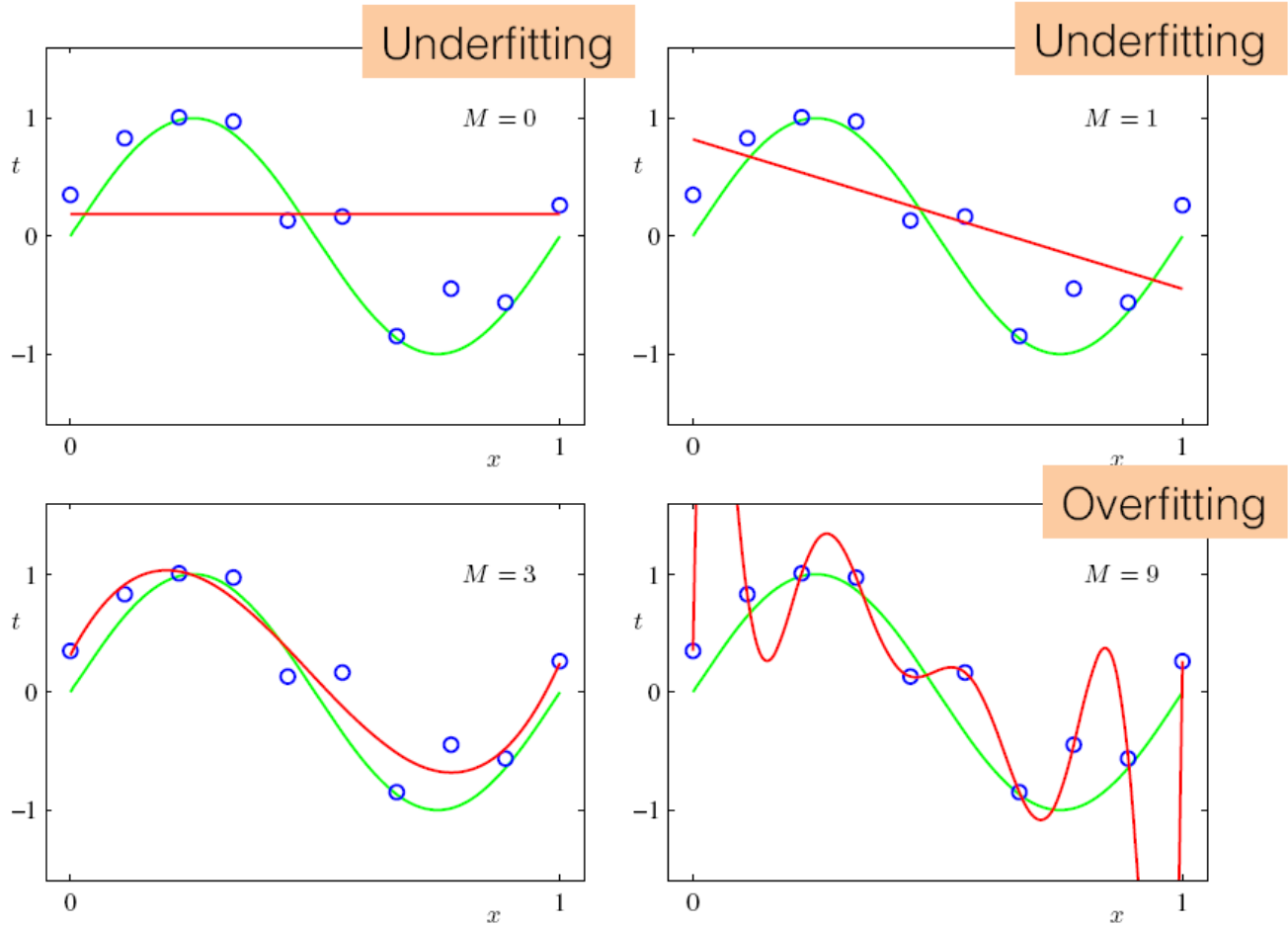
Critères des moindres carrés (MSE)

$$E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N \{y(x_n, \mathbf{w}) - t_n\}^2$$

Solution directe

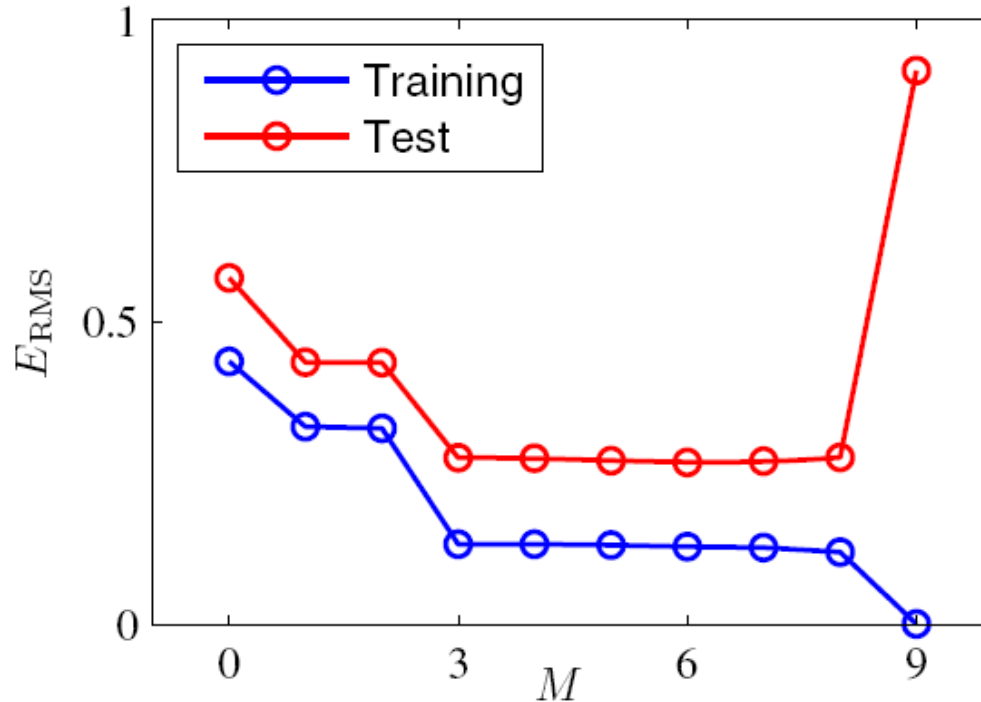


Choisir l'ordre M pour le polynôme?



Généralement on sépare les données en deux parties :

- Base d'apprentissage (*Training set*),
- Base de test (*test set*),



1. Re-substitution

Si toutes les données sont utilisées pour entraîner le modèle et que le taux d'erreur est évalué en fonction du résultat par rapport à la valeur réelle du même ensemble de données d'apprentissage, cette erreur est appelée erreur de resubstitution. Cette technique est appelée technique de validation de resubstitution.

2. Holdout

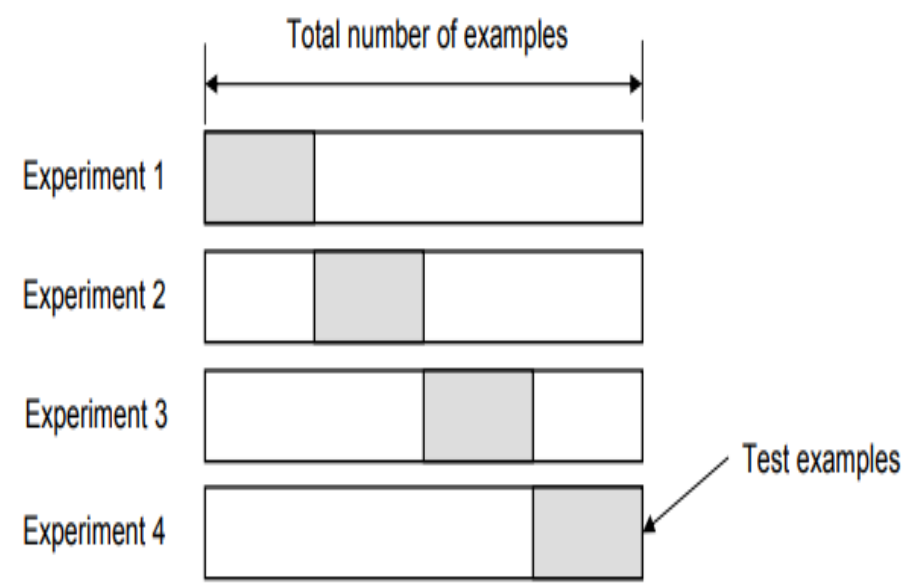
Pour éviter l'erreur de resubstitution, les données sont divisées en deux ensembles de données différents étiquetés comme un ensemble de données d'apprentissage et de test. Cela peut être une répartition 60/40 ou 70/30 ou 80/20. Cette technique est appelée technique de validation d'attente. Dans ce cas, il est probable qu'une distribution inégale de différentes classes de données soit trouvée dans l'ensemble de données d'apprentissage et de test. Pour résoudre ce problème, l'ensemble de données d'entraînement et de test est créé avec une répartition égale de différentes classes de données. Ce processus est appelé **stratification**.

Types de validation (1)

3. Validation croisée.....**la plus utilisée!!**

- La séparation des données en deux parties change itérativement.
- La mesure de performance est la moyenne sur toutes les itérations.

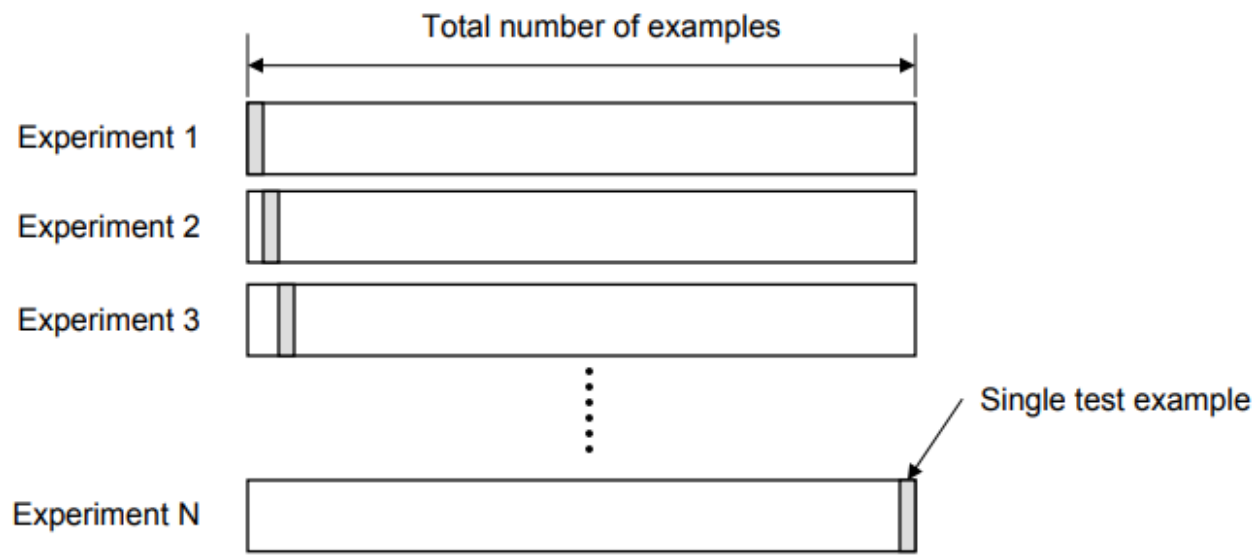
L'avantage est que des données entières sont utilisées pour l'apprentissage et le test. Le taux d'erreur du modèle est la moyenne des taux d'erreurs de différentes itérations. Cette technique peut également être appelée une forme de la méthode de rétention (hold-out) répétée. Le taux d'erreur pourrait être amélioré en utilisant la technique de stratification.



Types de validation (2)

4. Leave-One-Out Cross-Validation (LOOCV)

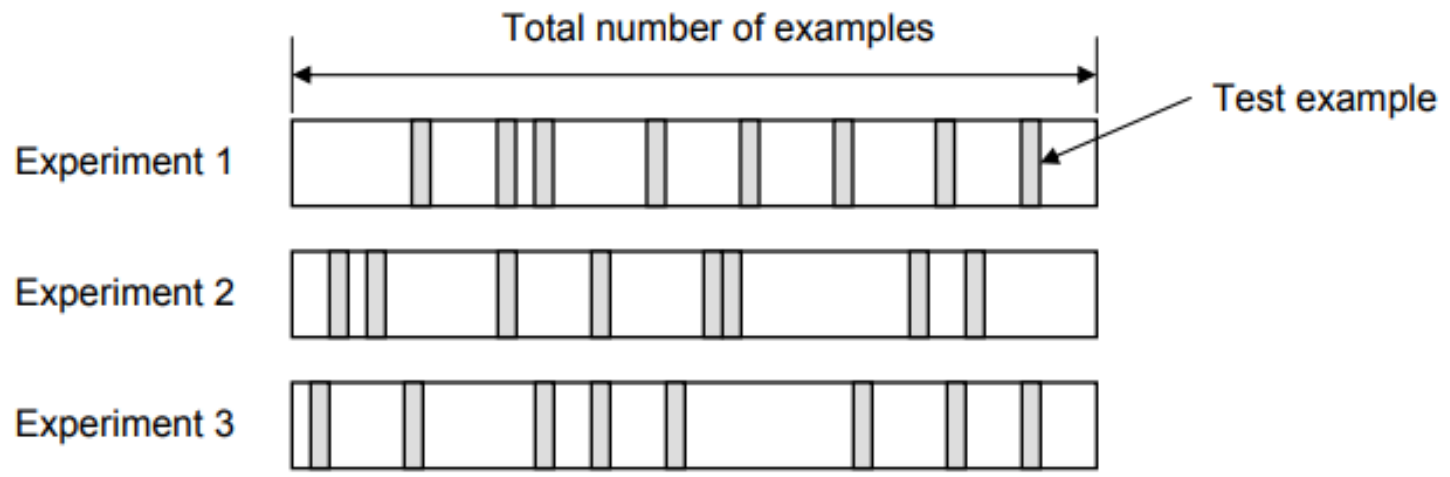
Dans cette technique, toutes les données, à l'exception d'une instance, sont utilisées pour l'apprentissage et une instance est utilisée pour le test. Ce processus est répété N fois s'il y a N instances. L'avantage est que des données entières sont utilisées pour les apprentissage et les tests. Le taux d'erreur du modèle est la moyenne du taux d'erreur de chaque itération.



Types de validation (3)

5. Random Subsampling

Dans cette technique, plusieurs ensembles de données sont choisis au hasard dans l'ensemble de données et combinés pour former un ensemble de données de test. Les données restantes forment l'ensemble de données d'apprentissage. Le diagramme suivant représente la technique de validation du sous-échantillonnage aléatoire. Le taux d'erreur du modèle est la moyenne du taux d'erreur de chaque itération.



6. Bootstrapping

Dans cette technique, l'ensemble de données d'apprentissage est sélectionné au hasard avec remplacement. Les exemples restants qui n'ont pas été sélectionnés pour l'apprentissage sont utilisés pour les tests. Contrairement à la validation croisée K-fold, la valeur est susceptible de changer d'une expérience à l'autre. Le taux d'erreur du modèle est la moyenne du taux d'erreur de chaque itération.

