**Cross validation**

<https://en.wikipedia.org/wiki/Cross-validation_(statistics)>

Exhaustive cross-validation: Exhaustive cross-validation methods are cross-validation methods which learn and test on all possible ways to divide the original sample into a training and a validation set.

<https://openclassrooms.com/fr/courses/4297211-evaluez-et-ameliorez-les-performances-dun-modele-de-machine-learning/4308241-mettez-en-place-un-cadre-de-validation-croisee>

* Il ne faut **jamais** évaluer un modèle sur des points qui ont été utilisés pour l’entraîner.
* On sépare donc les données entre **un jeu d’entraînement**, sur lequel on apprend le modèle, et **un jeu de test**, sur lequel on l’évalue.
* Pour utiliser l’intégralité de nos données pour entraîner et pour tester, et pour éviter un biais potentiel lié au fait de faire une évaluation unique, **on préfère faire une validation croisée**.
* Dans le cas d’un problème de classification, on fait attention à **stratifier la validation croisée** pour éviter d’introduire des biais.

**Leave-one-out**

Comment choisir le nombre de folds, k ?

Rappelez-vous, moins il y a de données disponibles pour l’entraînement, moins on est capable de bien apprendre. Or chaque fold contient (k−1)nk(k−1)nk  points (si nn est la taille du jeu complet). Si on pousse ce raisonnement, on fait autant de folds que de points dans le jeu complet (c’est à dire n )) , et les jeux d’entraînement font quasiment la même taille que le jeu complet ! C’est ce qu’on appelle le ***leave-one-out*** (on ne laisse de côté qu’un seul exemple pour chaque jeu d’entraînement).

Ça a l’air bien ! Sauf que…

* on augmente fortement le temps de calcul. Imaginez ça sur une base de données de 100 000 images ! Il faudrait entraîner 100 000 modèles sur 99 999 images chacun.
* on forme ainsi des jeux d’entraînements très similaires entre eux, et des jeux de test très différents les uns des autres. On va avoir quasiment le même modèle sur chaque fold, et la qualité de ses prédictions risque de beaucoup varier. Il sera difficile de tirer des conclusions.

En pratique, on choisit le plus souvent k=5 ou k=10.

<http://josephsalmon.eu/enseignement/Montpellier/Apprentissage_Statistique/CrossValidation_slides.pdf>