

□ **Random forest** – C’est une technique à base d’arbres qui utilise un très grand nombre d’arbres de décisions construits à partir d’ensembles de caractéristiques aléatoirement sélectionnés. Contrairement à un simple arbre de décision, il n’est pas interprétable du tout mais le fait qu’il ait une bonne performance en fait un algorithme populaire.

Remarque : les random forests sont un type de méthode ensembliste.

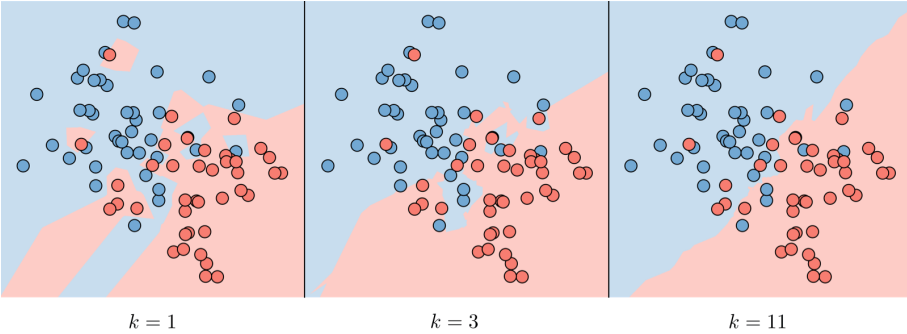
□ **Boosting** – L’idée des méthodes de boosting est de combiner plusieurs modèles faibles pour former un modèle meilleur. Les principales méthodes de boosting sont récapitulées dans le tableau ci-dessous :

Boosting adaptatif	Boosting par gradient
<ul style="list-style-type: none">- De grands coefficients sont mis sur les erreurs pour s’améliorer à la prochaine étape de boosting- Connue sous le nom d’Adaboost	<ul style="list-style-type: none">- Les modèles faibles sont entraînés sur les erreurs résiduelles

1.6 Autres approches non-paramétriques

□ **k -nearest neighbors** – L’algorithme des k plus proches voisins (en anglais *k-nearest neighbors*), aussi connu sous le nom de k -NN, est une approche non-paramétrique où la réponse d’un point est déterminée par la nature de ses k voisins du jeu de données d’entraînement. Il peut être utilisé dans des cadres de classification et de régression.

Remarque : Plus le paramètre k est élevé, plus le biais est élevé, et plus le paramètre k est faible, plus la variance est élevée.



1.7 Théorie d’apprentissage

□ **Inégalité de Boole** – Soit A_1, \dots, A_k k évènements. On a :

$$P(A_1 \cup \dots \cup A_k) \leq P(A_1) + \dots + P(A_k)$$

