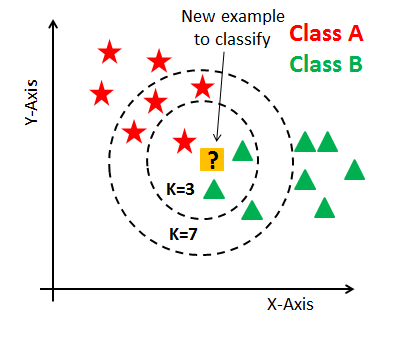
KNN, Regression Logisitique et validation

En guise de production, nous vous demandons de complétez la légende avec vos propres mots pour chacune des figures ci-dessous :

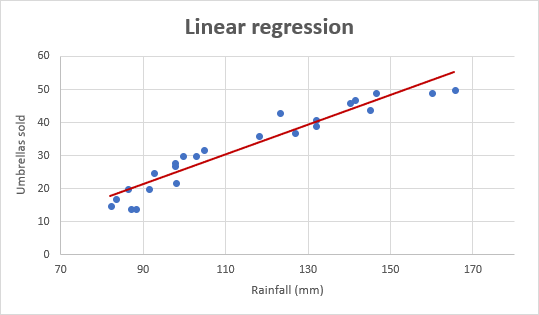
# KNN



Dans le classificateur KNN lorsque K=1 on classe un nouvel exemple en fonction de l’élément le plus proche donc ici les verts donc la classe B.

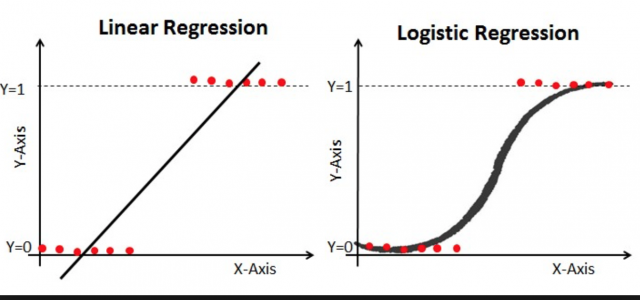
Lorsque K=3 on classe un nouvel exemple en fonction des exemples qui sont les plus proches et en fonction de la majorité ici c’est donc la classe B.

# Regression Linéaire



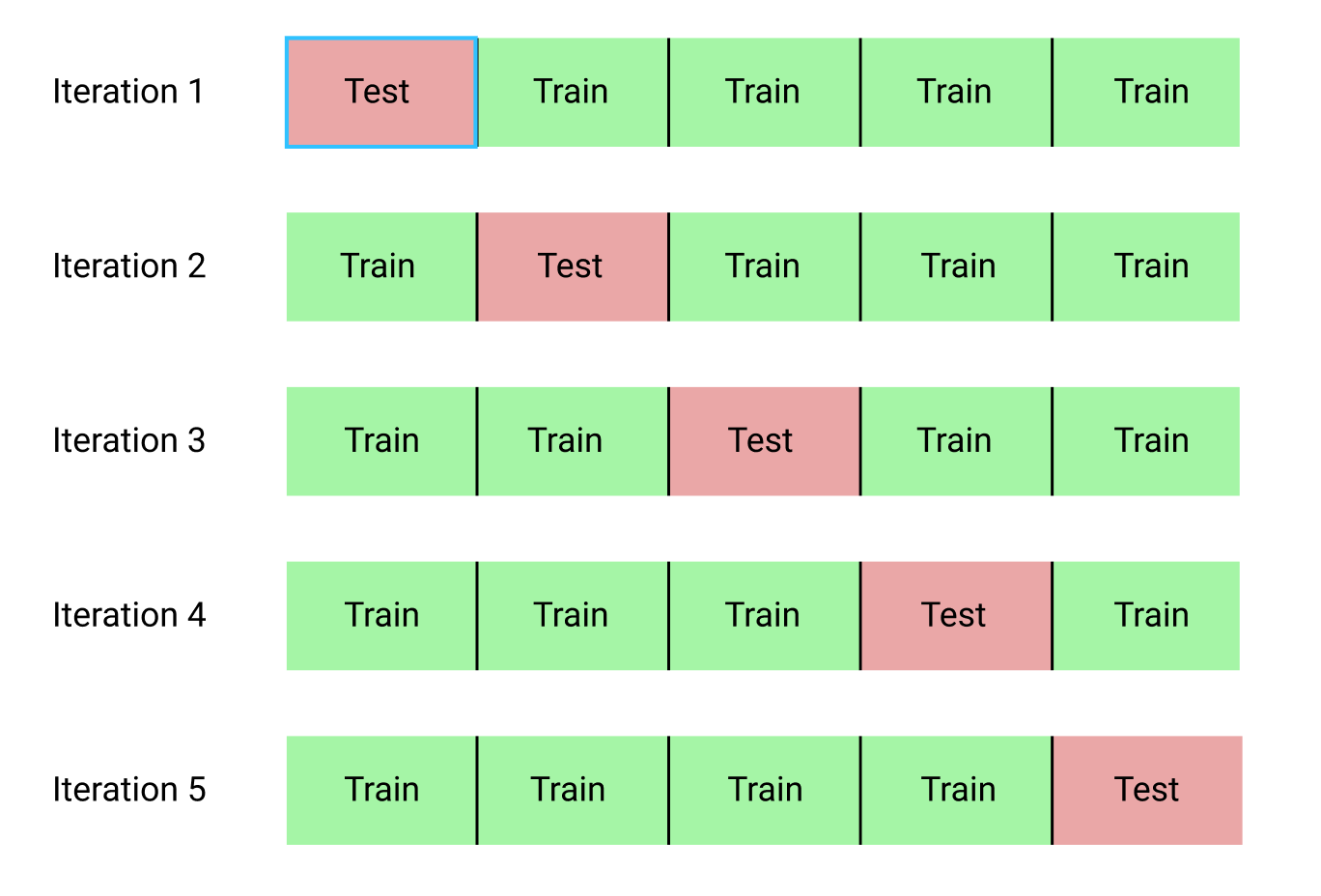
La régression linéaire prédit les valeurs de y en fonction de chaque valeur de x qui vont toutes se retrouver autour de la ligne rouge. Ici c’est donc en fonction du nombre de parapluies achetés qu’on va déterminer le nombre de millimètres de pluie qui est tombé ou inversement.

# Régression Logistisque



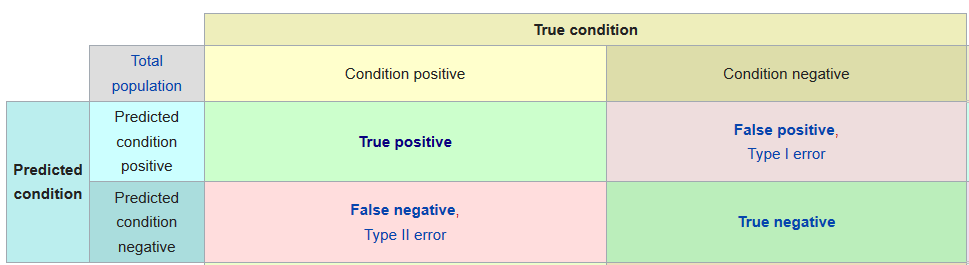
La régression logistique classe en utilisant la régression linéaire.

# Cross-validation N-Folds



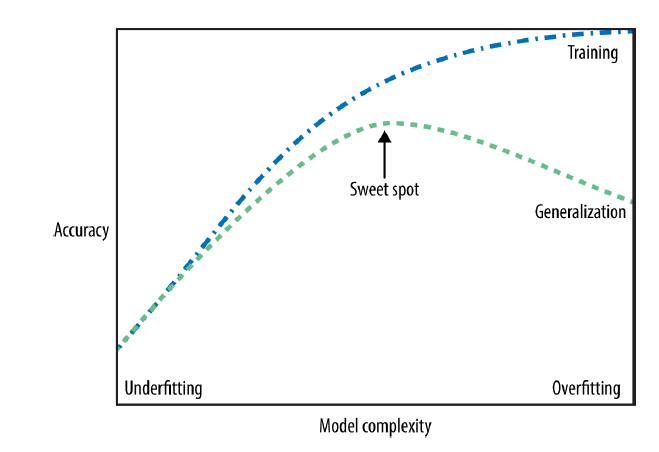
On utilise la cross-validation pour maximiser les performances et avoir un ratio qui s’approche le plus possible de 1 donc c'est à dire on divise en un certain nombre k le set de data et ensuite pour chaque K on fait une itération en testant le training set avec le test set. Et on compare ensuite pour trouver le meilleur training set.

# True and False Positives



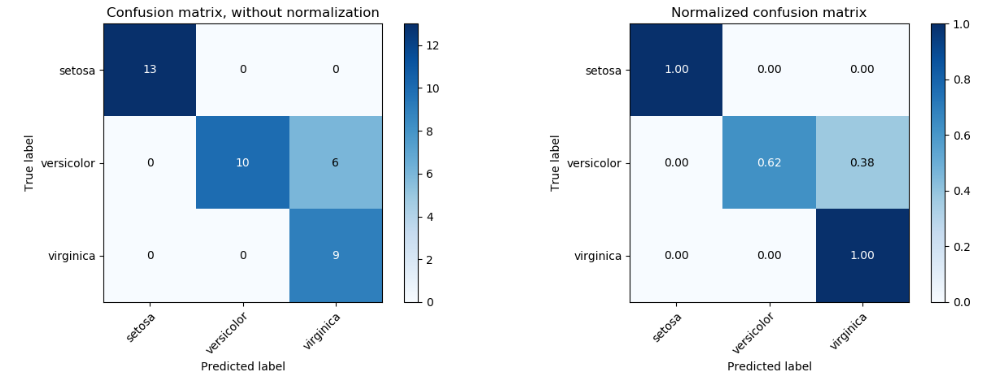
Un faux positif c'est lorsque le classificateur prédit que le test va être positif alors que la condition est négative.

# Over/underfitting



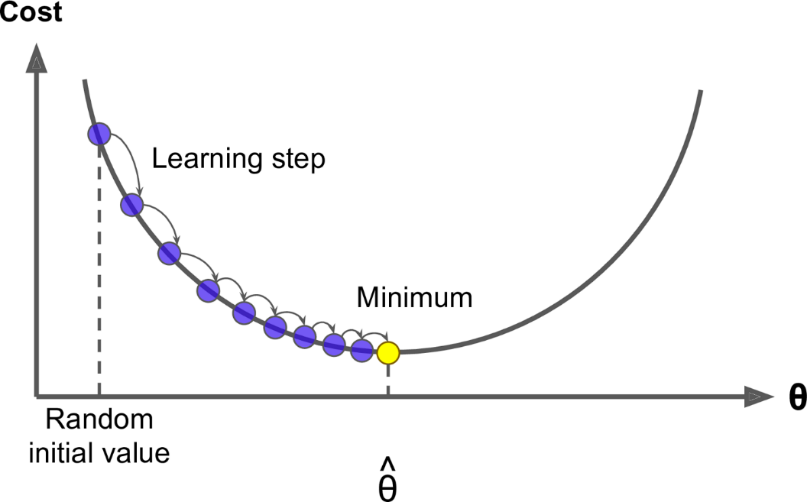
Il y a overfitting lorsque le model est trop basé sur les données d’entrainements et ne généralise pas assez.

# Matrice de confusion



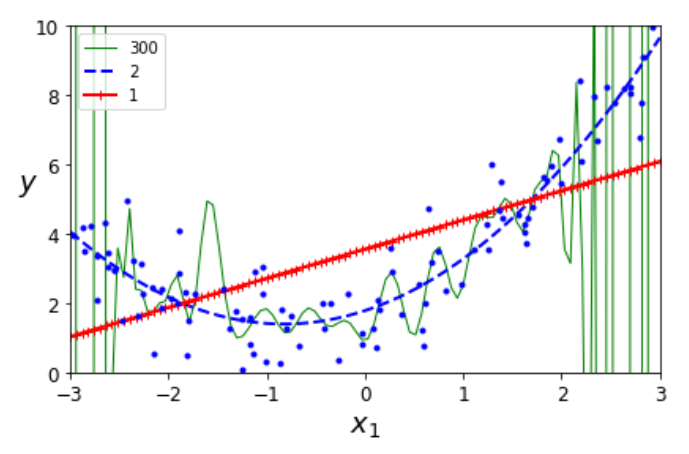
Matrice de confusion permet de voir avec des chiffres les moments ou l'algorithme s'est trompé et où ça. Par exemple, ici sur le graphe le label recherché était versicolor mais pourtant l’IA a prédit virginica et on voit qu'il s'est trompé 6 fois. Les chiffres en diagonale du haut à gauche au bas à droite sont les parties réussies.

# Descente de gradient



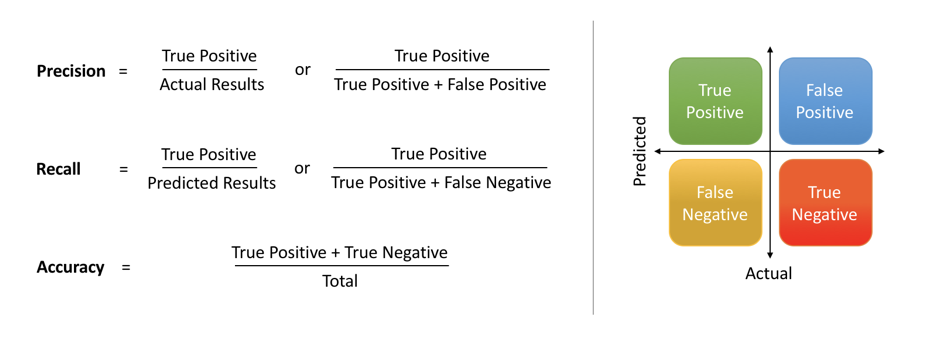
La descente de gradient est utilisée dans la régression linéaire (et dans beaucoup d’autres méthodes) pour minimiser l’équation d’erreur de la MSE et on essaye de trouver un jeu de paramètres pour celui-ci. Il essaye de trouver le minimum en faisant des essais erreurs.

Régression polynomiale



La courbe rouge représente une régression linéaire, la bleu une régression polynomiale.

# Precision/Recall



Intuitivement le recall mesure le pourcentage de résultats positifs que l’intelligence artificielle a bien prédit tandis que la précision mesure le pourcentage de résultats positifs sur les résultats positifs qu’ils soient bon ou pas.