

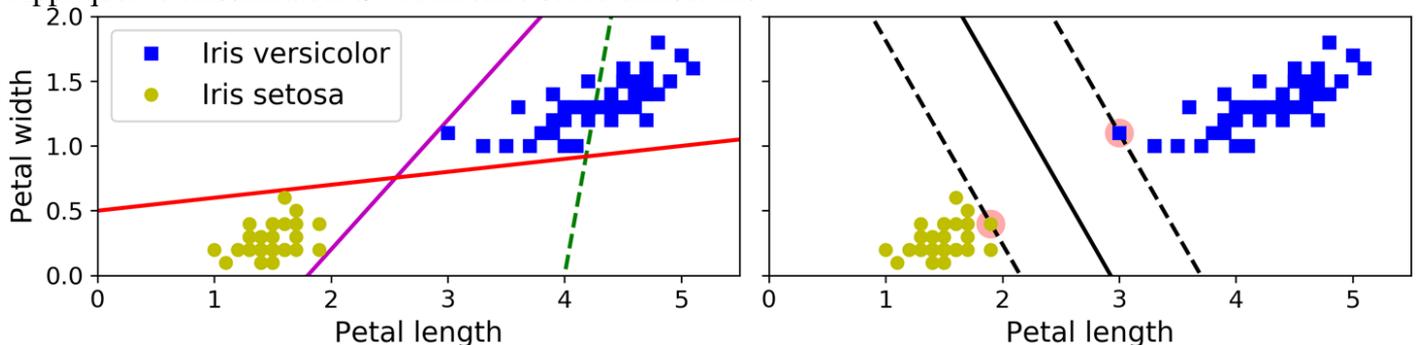
TP 6 : Machine à vecteurs de supports

1. Classification SVM linéaire :

Vous pouvez considérer qu'un classificateur SVM correspond à la rue la plus large possible (représentée par les lignes pointillées parallèles) entre les classes. C'est ce qu'on appelle la classification à grande marge.

Question:

Appliquer le classificateur SVM linéaire sur le dataset Iris.



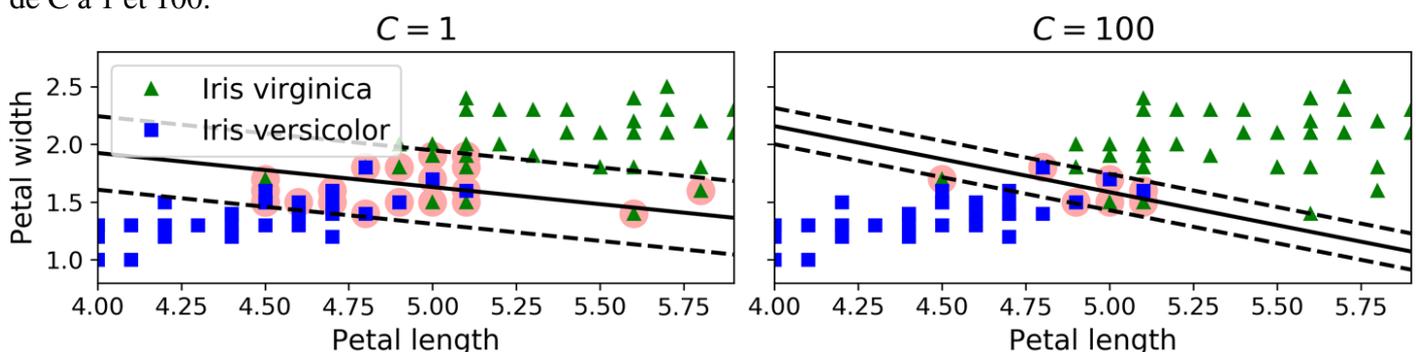
2. Cas non séparable de SVM :

Pour éviter les problèmes des cas non séparables de SVM, utilisez un modèle plus flexible. L'objectif est de trouver un bon équilibre entre garder la rue aussi large que possible et limiter les violations de marge (c'est-à-dire les instances qui se retrouvent au milieu de la rue ou même du mauvais côté). C'est ce qu'on appelle la classification des marges souples.

Lors de la création d'un modèle SVM à l'aide de Scikit-Learn, nous pouvons spécifier le nombre d'hyperparamètres. C'est l'un de ces hyperparamètres.

Question:

Appliquer le classificateur SVM linéaire pour le cas non séparable sur le dataset Iris en choisissant les valeurs de C à 1 et 100.

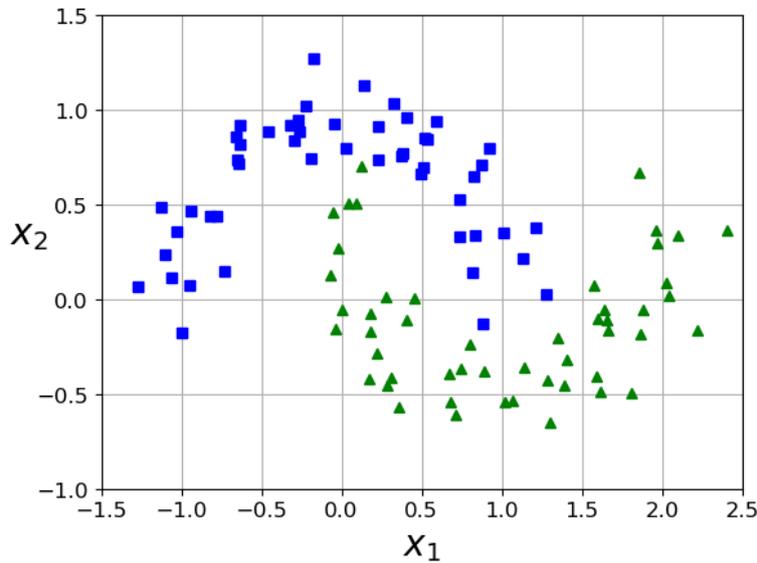


3. Classification SVM non linéaire :

Malgré que les classificateurs SVM linéaires soient efficaces et fonctionnent très bien dans de nombreux cas, de nombreux datasets ne sont pas linéairement séparables. Une approche pour le traitement des datasets non linéaires consiste à générer des attributs, telles que des attributs polynomiales; dans certains cas, cela peut aboutir à un dataset linéairement séparables.

Question:

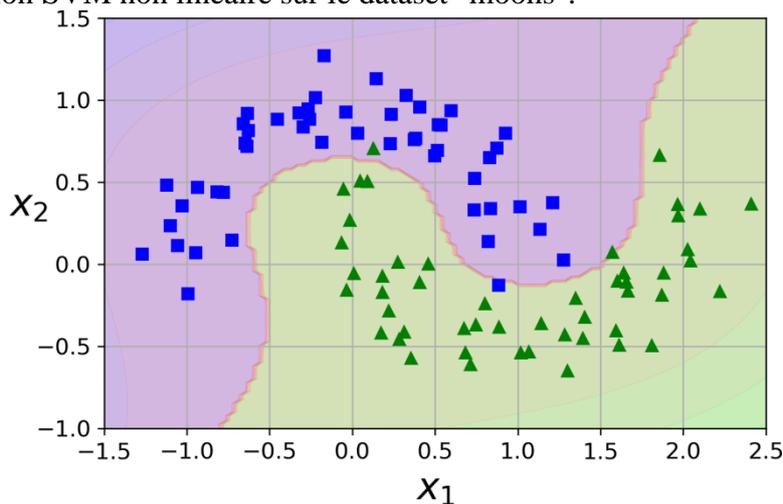
Générer l'ensemble de données (dataset) moons. Afficher ces données à l'aide d'un graphe 2D.



Pour implémenter cette idée à l'aide de Scikit-Learn, créez un pipeline contenant un transformateur PolynomialFeatures, suivi d'un StandardScaler et d'un LinearSVC. Testons cela sur le dataset qui s'appelle "moons": il s'agit d'un ensemble de données fictives proposé pour la classification binaire dans lequel les points de données ont la forme de deux demi-cercles entrelacés. Vous pouvez générer cet ensemble de données à l'aide de la fonction `make_moons()`.

Question:

Appliquer la classification SVM non linéaire sur le dataset "moons".



4. Noyau polynomial :

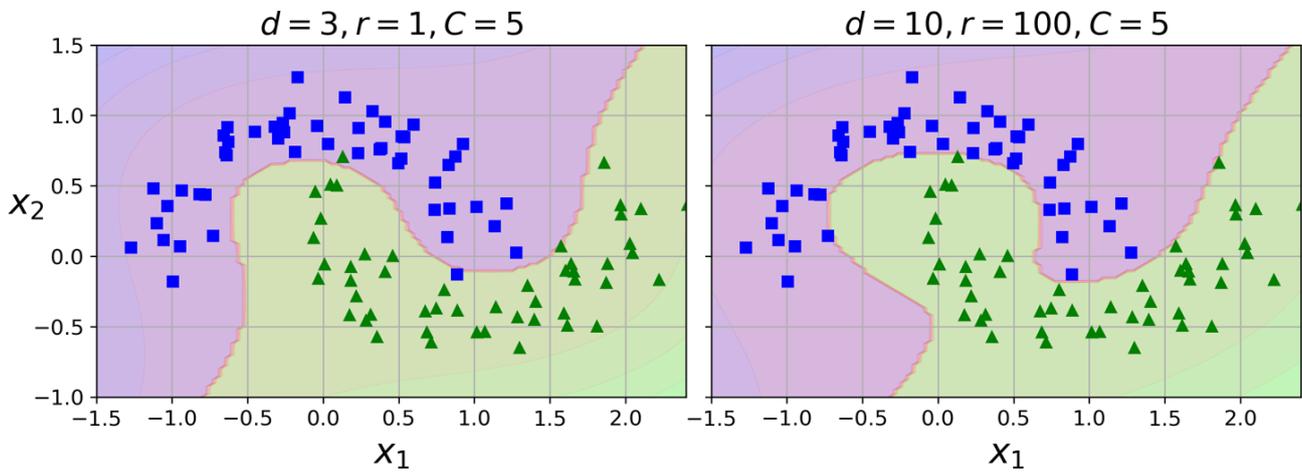
L'astuce du noyau permet d'obtenir le même résultat que si vous aviez ajouté de nombreux attributs polynomiaux, même avec des polynômes de très haut degré, sans avoir à les ajouter. Il n'y a donc pas d'explosion combinatoire du nombre d'attributs car vous n'ajoutez aucuns attributs. La classe SVC implémente cette astuce.

Question:

Tester l'astuce du noyau polynomila sur le dataset Moons pour deux valeurs de degrés 3 et 10.

Note:

Si votre modèle souffre du sur-apprentissage, vous voudrez peut-être réduire le degré polynomial. Inversement, s'il est sous- apprentissage, vous pouvez essayer de l'augmenter. L'hyperparamètre `coef0` contrôle la mesure dans laquelle le modèle est influencé par les polynômes de haut degré par rapport aux polynômes de bas degré.



5. Noyau RBF gaussien :

Tout comme l'astuce des attributs polynomiales, la méthode des attributs de similarité peut être utile avec n'importe quel algorithme d'apprentissage automatique, mais il peut être coûteux en calcul de calculer toutes les attributs supplémentaires, en particulier sur de grands ensembles d'apprentissage. Encore une fois, l'astuce du noyau fait sa magie SVM, permettant d'obtenir un résultat similaire comme si vous aviez ajouté de nombreux attributs de similarité.

Question:

Appliquer la classification SVM avec le noyau RBF gaussien.

