

핸즈온 머신러닝 4주차 예습

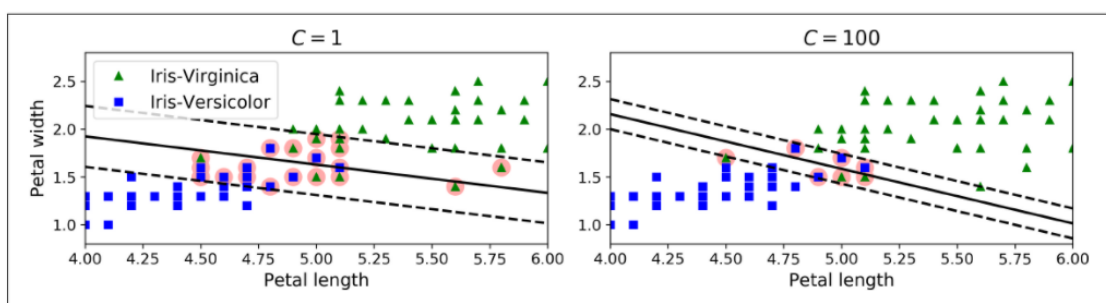
서포트 벡터 머신 SVM : 복잡한 분류 문제에 적합 (스케일링 권장)

클래스 사이에 가장 폭이 넓은 도로를 찾아 : Large margin classification

C: 하이퍼 파라미터 지정

ex. 과대적합의 경우 C를 감소시켜 모델을 규제

▼ 선형 SVM 분류



```
import numpy as np
from sklearn import datasets
from sklearn.pipeline import Pipeline
from sklearn.preprocessing import StandardScaler
from sklearn.svm import LinearSVC

iris = datasets.load_iris()
X = iris["data"][:, (2, 3)] # 꽃잎 길이, 꽃잎 너비
y = (iris["target"] == 2).astype(np.float64) # Iris virginica

svm_clf = Pipeline([
    ("scaler", StandardScaler()),
    ("linear_svc", LinearSVC(C=1, loss="hinge", random_state=42)),
])

svm_clf.fit(X, y)
```

```
svm_clf.predict([[5.5, 1.7]]) #array([1.])
```

▼ 비선형 SVM 분류

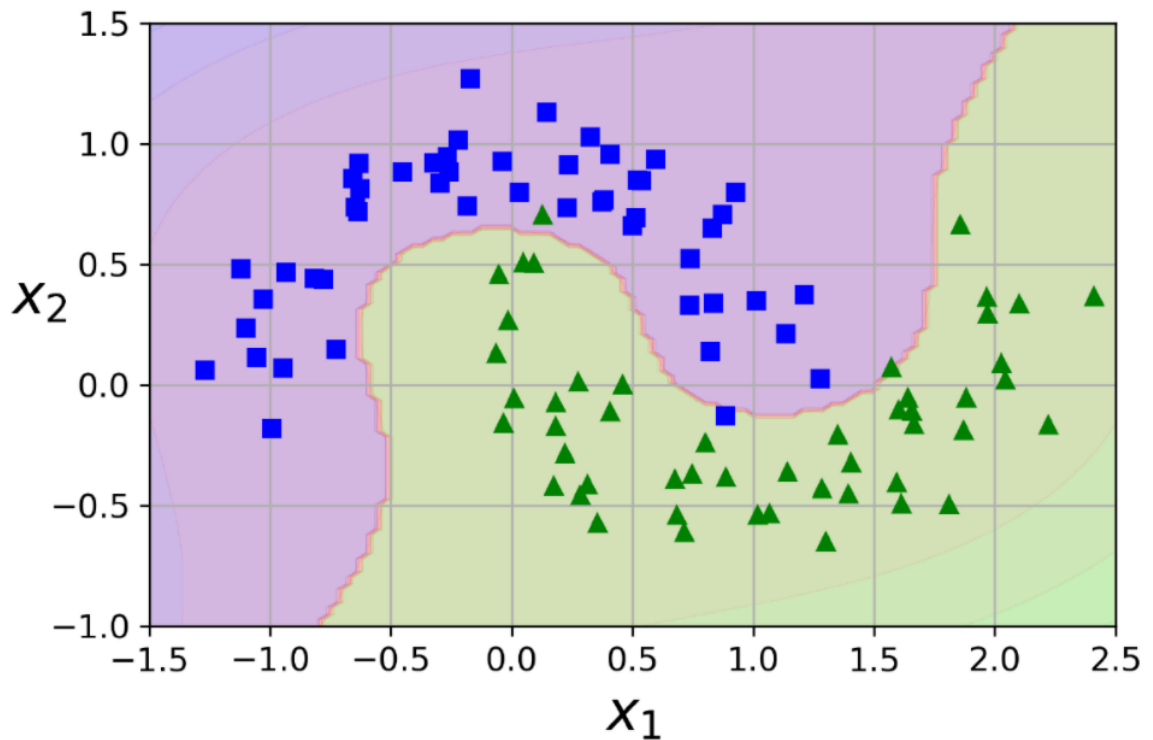
```
from sklearn.datasets import make_moons
from sklearn.pipeline import Pipeline
from sklearn.preprocessing import PolynomialFeatures
```

```

polynomial_svm_clf = Pipeline([
    ("poly_features", PolynomialFeatures(degree=3)),
    ("scaler", StandardScaler()),
    ("svm_clf", LinearSVC(C=10, loss="hinge", random_state=42))
])

polynomial_svm_clf.fit(X, y)

```



다항식 커널

- 커널 트릭 : 실제로는 특성을 추가하지 않으면서 다항식 특성을 많이 추가한 것과 같은 결과

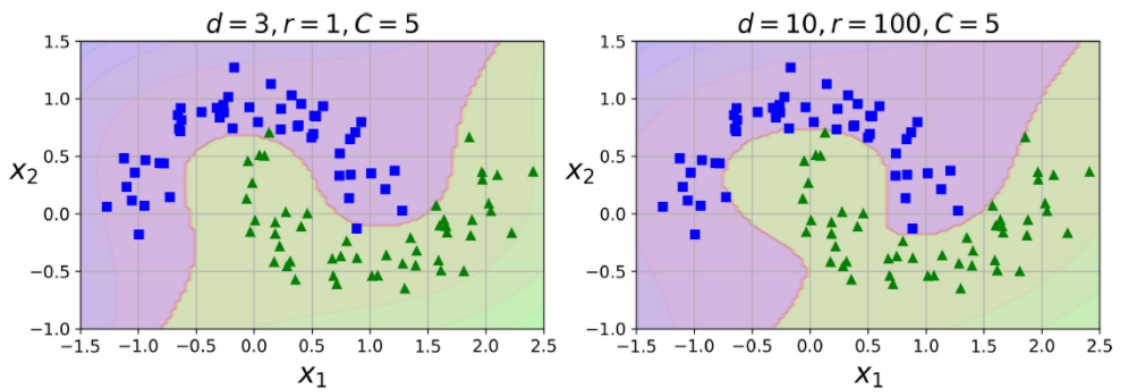
```

from sklearn.svm import SVC

poly_kernel_svm_clf = Pipeline([
    ("scaler", StandardScaler()),
    ("svm_clf", SVC(kernel="poly", degree=3, coef0=1, C=5))
])

poly_kernel_svm_clf.fit(X, y)

```



▼ SVM 회귀

제한된 마진(도로 밖의 샘플) 안에서 도로 안에 가능한 많은 샘플이 들어가도록 학습
→ 목표를 반대로 하는 것

```
from sklearn.svm import LinearSVR

svm_reg = LinearSVR(epsilon=1.5, random_state=42)
svm_reg.fit(X, y)
```