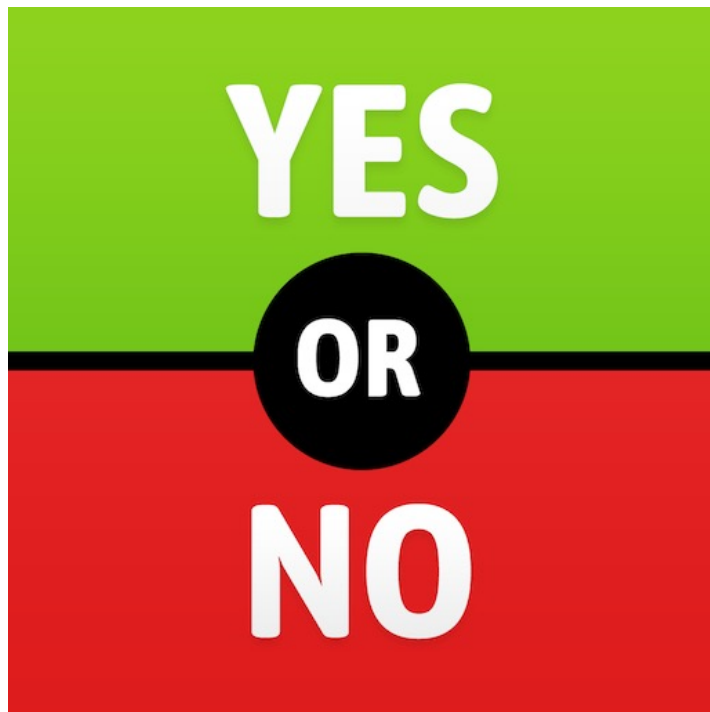


CUAI 스터디_MTs (SVM)

2022.05.01(일)

발표자 : 최윤한

이진분류란? (Binary Classification)



SVM(Support Vector Machine)

Decision Boundary(결정경계)를
결정하기 위한 이진분류모델

=> Margin이 최대화 된
초평면(hyperplane)을 찾는 모델

Margin??

결정경계와 서포트벡터 까지의 거리

꽃잎 데이터

```
from sklearn import datasets
```

```
iris = datasets.load_iris()
```

```
from sklearn.svm import SVC  
from sklearn import datasets
```

```
iris = datasets.load_iris()  
X = iris["data"][:, (2, 3)] # 꽃잎 길이, 꽃잎 너비  
y = iris["target"]
```

```
setosa_or_versicolor = (y == 0) | (y == 1)  
X = X[setosa_or_versicolor]  
y = y[setosa_or_versicolor]
```



caseno	일련번호이다. (1부터 150까지 입력된다.)
Sepal Length	꽃받침의 길이 정보이다.
Sepal Width	꽃받침의 너비 정보이다.
Petal Length	꽃잎의 길이 정보이다.
Petal Width	꽃잎의 너비 정보이다.
Species	꽃의 종류 정보이다. setosa / versicolor / virginica 의 3종류로 구분된다.

SVM(Support Vector Machine)

```
from sklearn import datasets

iris = datasets.load_iris()
```

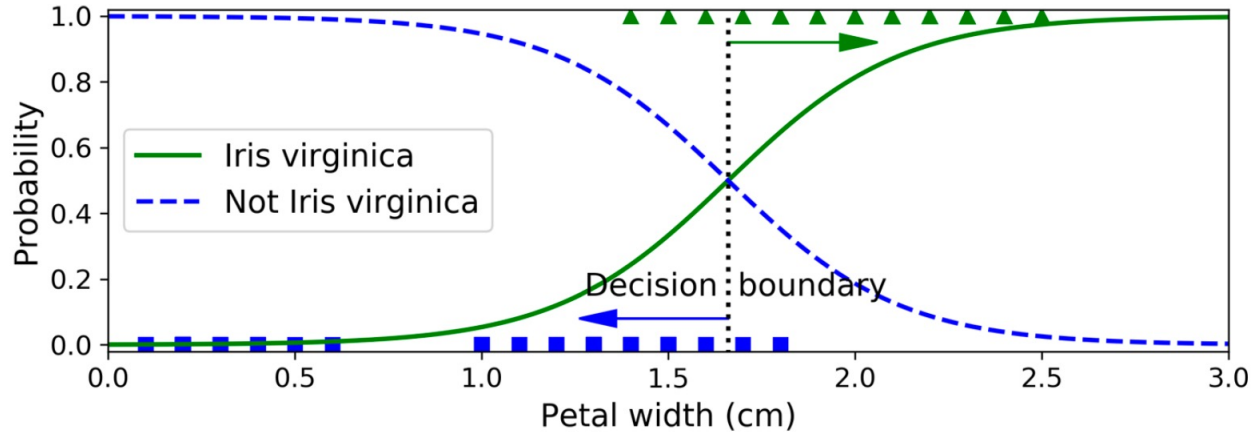
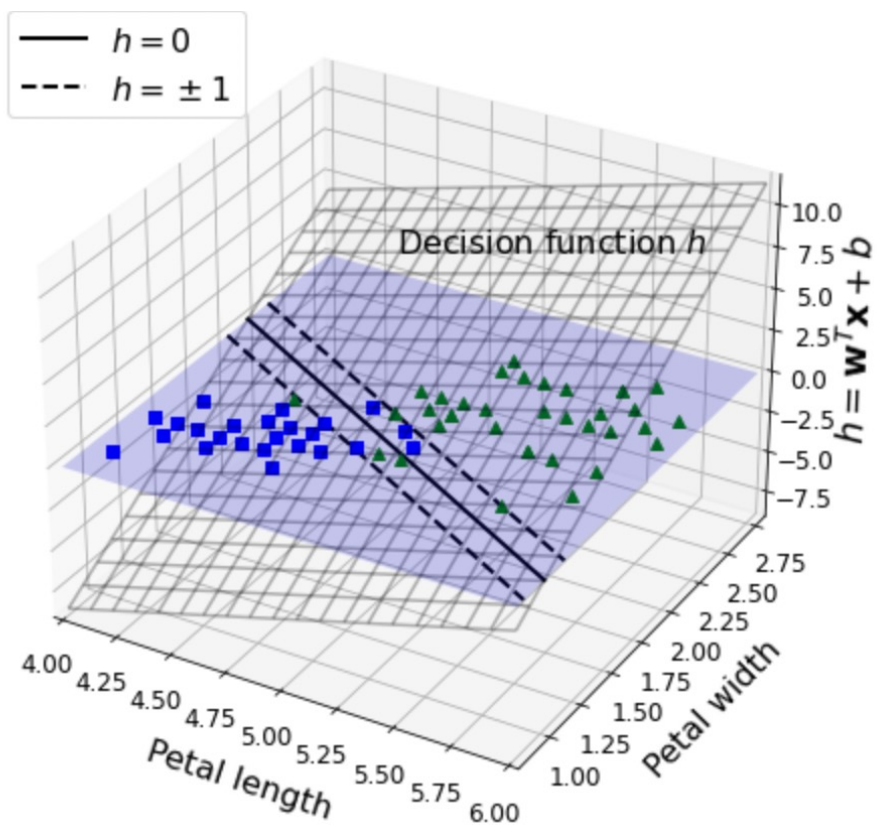


Figure 4-23. Estimated probabilities and decision boundary



SVM(Support Vector Machine)

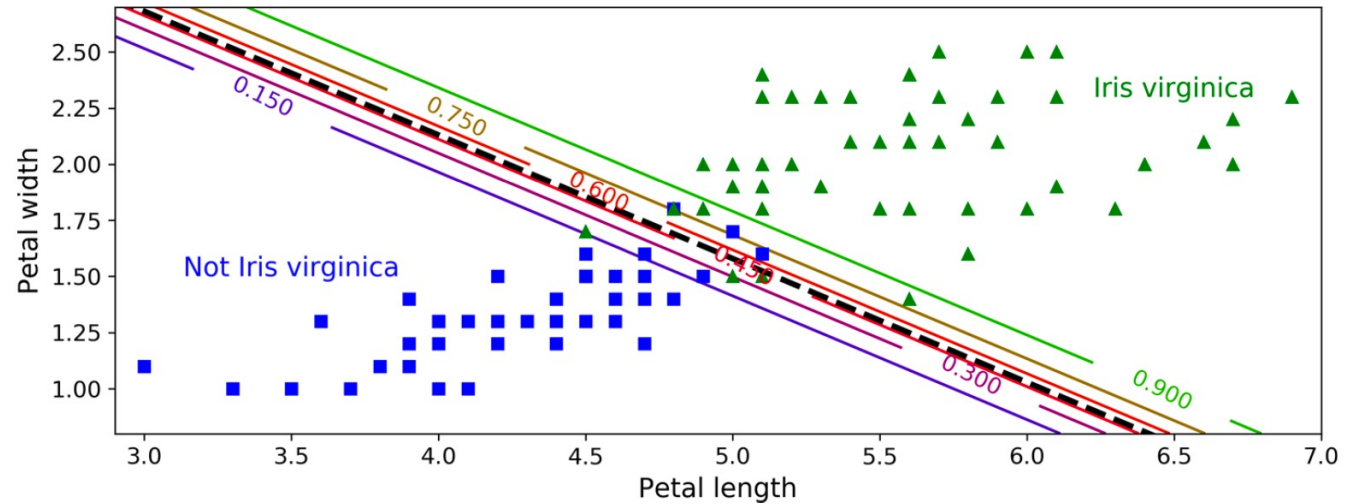


Figure 4-24. Linear decision boundary

SVM(Support Vector Machine)

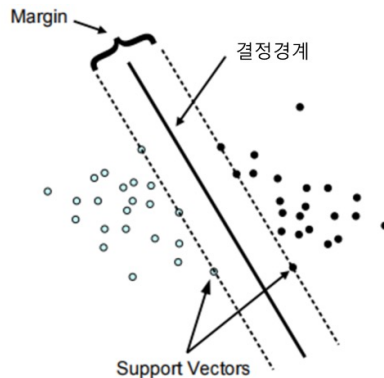
하드 마진 분류, 소프트 마진 분류

Hard margin Classification

- linear separable 할 때, 잘 작동한다
- support vector를 잘 골라내야하기 때문에 이상치에 민감하다.
- margin 값을 작게 잡아, 개별적인 데이터에 대해 신경쓴다. \Rightarrow overfitting의 문제가 생긴다.

Soft margin Classification

- margin 값을 크게 잡아, decision boundary와 서포트벡터들이 멀어진다. \Rightarrow under fitting의 문제



$C(\text{cost})$

SVM(Support Vector Machine)

```
from sklearn.datasets import make_moons
X, y = make_moons(n_samples=100, noise=0.15, random_state=42)

def plot_dataset(X, y, axes):
    plt.plot(X[:, 0][y==0], X[:, 1][y==0], "bs")
    plt.plot(X[:, 0][y==1], X[:, 1][y==1], "g^")
    plt.axis(axes)
    plt.grid(True, which='both')
    plt.xlabel(r"$x_1$", fontsize=20)
    plt.ylabel(r"$x_2$", fontsize=20, rotation=0)

plot_dataset(X, y, [-1.5, 2.5, -1, 1.5])
plt.show()
```

SVM(Support Vector Machine) - 다항식 커널

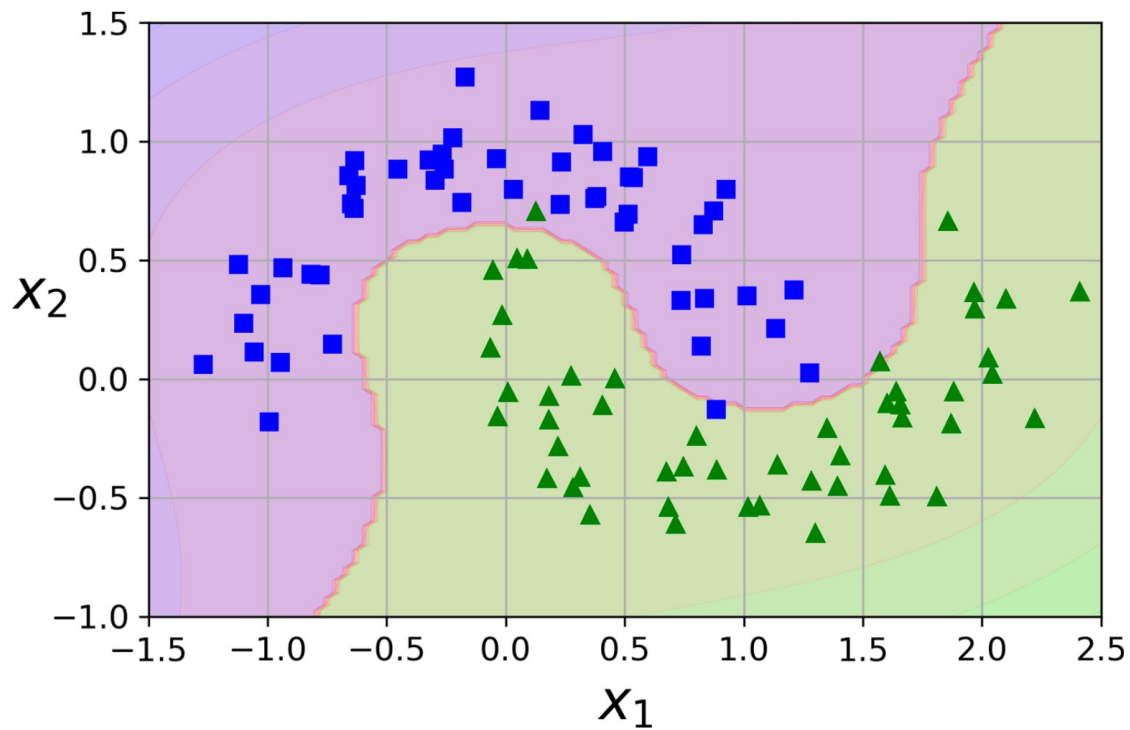
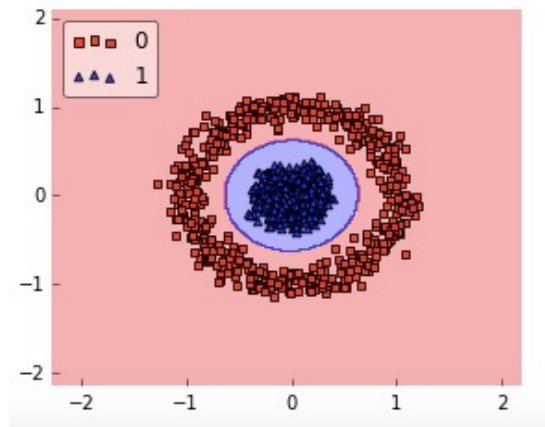
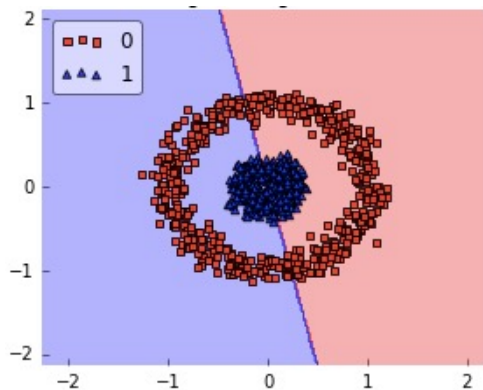


Figure 5-6. Linear SVM classifier using polynomial features

SVM(Support Vector Machine)

RBF 커널(가우시안 커널)



SVM(Support Vector Machine)

RBF 커널(가우시안 커널)

Kernel

- SVM에서 목적함수와 예측 모델을 dual form으로 변형가능하다.
 - 두개의 변환된 독립변수 벡터를 inner product한 값을 하나의 함수로 나타낼 수 있다.

$$k(x_i, x_j) = \phi(x_i)^T \phi(x_j)$$

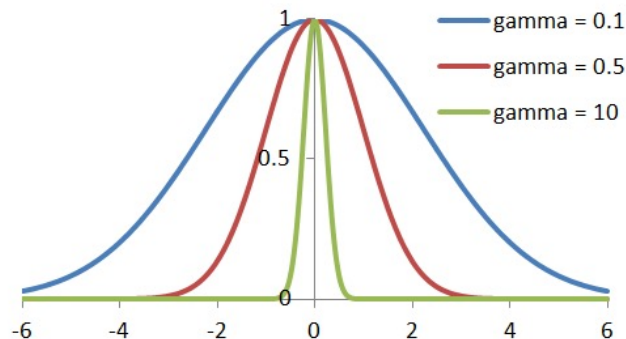
저차원 => 고차원

SVM(Support Vector Machine)

RBF 커널(가우시안 커널)

Gamma??

영향력을 행사하는 거리 -> 규제
역할(가우시안 함수에서
표준편차와 관련되어 있다.)



Gamma r값에 따른
정규분포 그래프

- Gamma가 클수록 규제가 심하다.
- > 각 데이터 샘플의 영향력이 높아진다.
- > 결정경계의 곡률을 결정한다.
- > 클수록 Overfitting

SVM(Support Vector Machine) 사용하기;

```
# SVM import
from sklearn.svm import SVC
from sklearn import datasets

# iris 꽃잎 데이터 불러오기
iris = datasets.load_iris()
X = iris["data"][:, (2, 3)] # 꽃잎 길이, 꽃잎 너비
y = iris["target"]

setosa_or_versicolor = (y == 0) | (y == 1)
X = X[setosa_or_versicolor]
y = y[setosa_or_versicolor]

# SVM 분류 모델
# kernel = "linear" => 선형 분류기
svm_clf = SVC(kernel="linear", C=float("inf"))

# 학습
svm_clf.fit(X, y)
```

SVM(Support Vector Machine) 사용하기

```
from sklearn.svm import SVC
```

```
from sklearn.svm import SVC

# 다항식 커널 파이프라인 생성
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(Support Vector Machine) 사용하기

```
from sklearn.svm import SVC
```

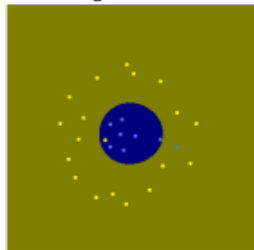
```
from sklearn.svm import SVC

gamma1, gamma2 = 0.1, 5
C1, C2 = 0.001, 1000
hyperparams = (gamma1, C1), (gamma1, C2), (gamma2, C1), (gamma2, C2)

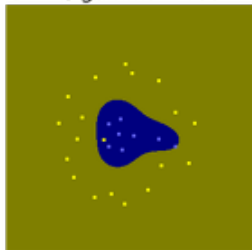
svm_clfs = []
for gamma, C in hyperparams:
    rbf_kernel_svm_clf = Pipeline([
        ("scaler", StandardScaler()),
        ("svm_clf", SVC(kernel="rbf", gamma=gamma, C=C))
    ])
    rbf_kernel_svm_clf.fit(X, y)
    svm_clfs.append(rbf_kernel_svm_clf)
```


gamma 커짐

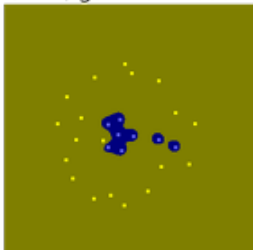
C = 1, gamma = 10



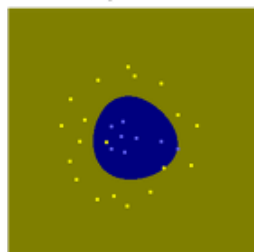
C = 1, gamma = 100



C = 1, gamma = 1000



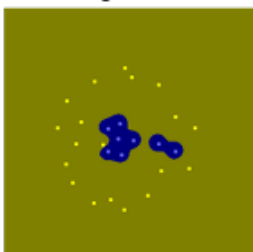
C = 10, gamma = 10



C = 10, gamma = 100



C = 10, gamma = 1000



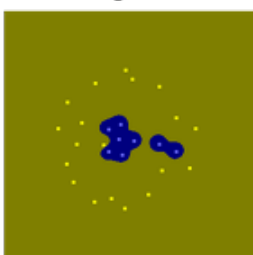
C = 100, gamma = 10



C = 100, gamma = 100



C = 100, gamma = 1000



C 커짐

SoftMax

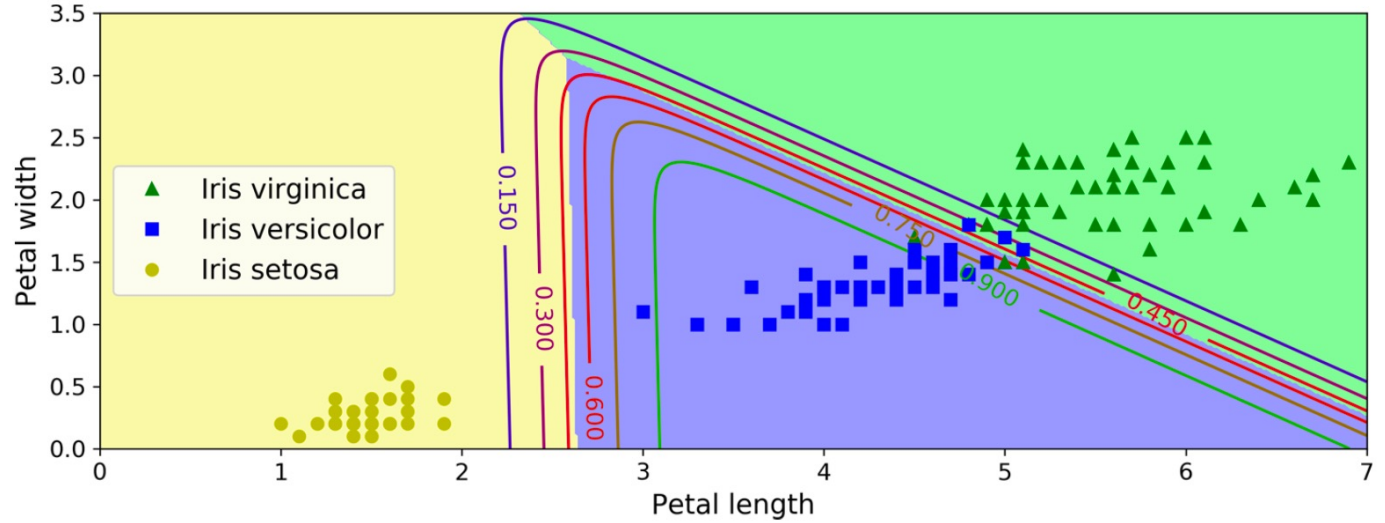


Figure 4-25. Softmax Regression decision boundaries

References

- <https://ai-times.tistory.com/418>
- <https://bskyvision.com/163>