

問題1. `ksvm{kernlab}`の三種類(線形、多項式、ガウシアン)カーネルを用いてデータ`BC.train`で学習し、`BC.test`の正解率を求め、どのカーネルの正解率が最も高いかを説明せよ。

データの作成

```
In [21]: from sklearn.datasets import load_breast_cancer
from sklearn.model_selection import train_test_split
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score
import pandas as pd
```

```
In [16]: data = load_breast_cancer()
X = data.data
df_X = pd.DataFrame(X)
df_X.columns = data.feature_names
Y = data.target
df_Y = pd.DataFrame(Y)
df_Y.columns = ["class"]
```

```
In [17]: X_train, X_test, y_train, y_test = train_test_split(df_X, df_Y, test_size = 0.3, random_state = 1)
```

各カーネルSVM (線形, 多項式, ガウシアン) のインスタンス化

```
In [22]: linear_svm = SVC(kernel = "linear")
poly_svm = SVC(kernel = "poly", degree = 8)
gaussian_svm = SVC(kernel = "rbf")
```

学習

```
In [24]: pred_linear = linear_svm.fit(X_train, y_train).predict(X_test)
acc_linear = accuracy_score(pred_linear, y_test)

pred_poly = poly_svm.fit(X_train, y_train).predict(X_test)
acc_poly = accuracy_score(pred_poly, y_test)

pred_gaussian = gaussian_svm.fit(X_train, y_train).predict(X_test)
acc_gaussian = accuracy_score(pred_gaussian, y_test)

print(f"線形カーネルを用いた正解率: {round(acc_linear, 3)}")
print(f"多項式カーネルを用いた正解率: {round(acc_poly, 3)}")
print(f"ガウシアンカーネルを用いた正解率: {round(acc_gaussian, 3)}")
```

```
C:\Users\MK209\Anaconda3\lib\site-packages\sklearn\utils\validation.py:72: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
    return f(**kwargs)
```

```
線形カーネルを用いた正解率: 0.953
多項式カーネルを用いた正解率: 0.889
ガウシアンカーネルを用いた正解率: 0.918
```

```
C:\Users\MK209\Anaconda3\lib\site-packages\sklearn\utils\validation.py:72: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
    return f(**kwargs)
```

```
C:\Users\MK209\Anaconda3\lib\site-packages\sklearn\utils\validation.py:72: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
    return f(**kwargs)
```

線形カーネルを用いたSVMの正解率が最も高い