Introduction to Machine Learning Homework 4 Announcement

Presenter: TA Jui-Che (Ben)

Lastest update: 2023/11/28 12:00

Homework 4

- Deadline: 23:59, Dec. 19th (Tue), 2023
- Coding (50%): Implement kernel methods by only using **numpy.**
  - Submit your python file (.py).
  - Answer the questions (by screenshots) in the report (.pdf).
- Handwritten Questions (50%): Answer questions about kernel methods.
  - Answer the questions (handwritten, typed, digital, etc.) in the report.

Links

- Questions: Link
- Sample code: Link
- Dataset: Link
- Report template: Link (same as HW1)

Environment

- Python version: 3.8 or newer
- Tips
  - We recommend that you use **virtual environments** when implementing your homework assignments.
  - Here are some popular virtual environment management tools:
    - Conda
    - Miniconda
    - virtualenv

Numpy

- Build-in array operations.
- Numpy Tutorial: [Link]

```
a = np.array([1, 2, 3])
b = np.array([4, 5, 6])
for i in range(a.shape[0]):
    a[i] *= b[i]
print(a)
# a = [ 4 10 18]
```

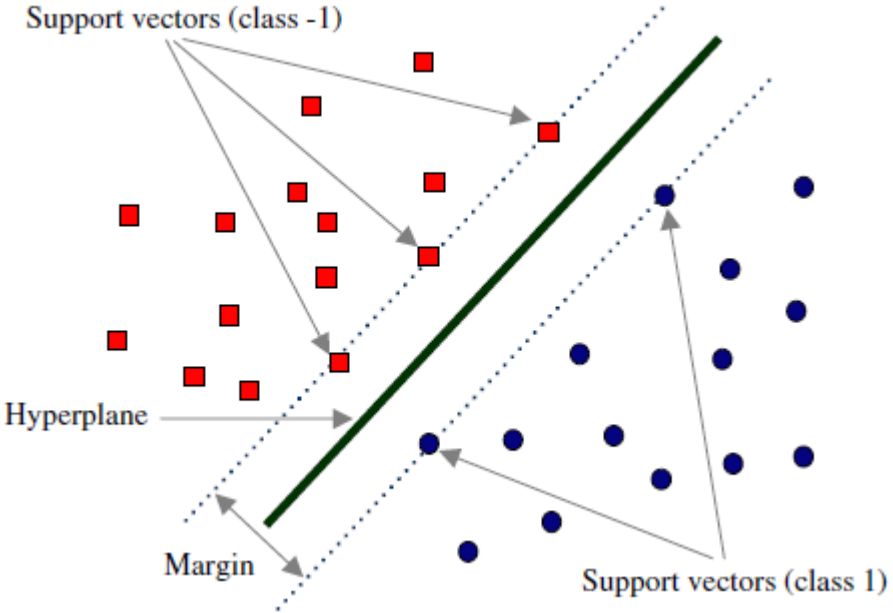
```
a = np.array([1, 2, 3])
b = np.array([4, 5, 6])
a *= b
print(a)
# a = [ 4 10 18]
```

```
import math
a = np.array([1, 4, 9])
for i in range(a.shape[0]):
    a[i] = math.sqrt(a[i])
print(a)
# a = [1 2 3]
```

```
a = np.array([1, 4, 9])
a = np.sqrt(a)
print(a)
# a = [1 2 3]
```

Support Vector Machine

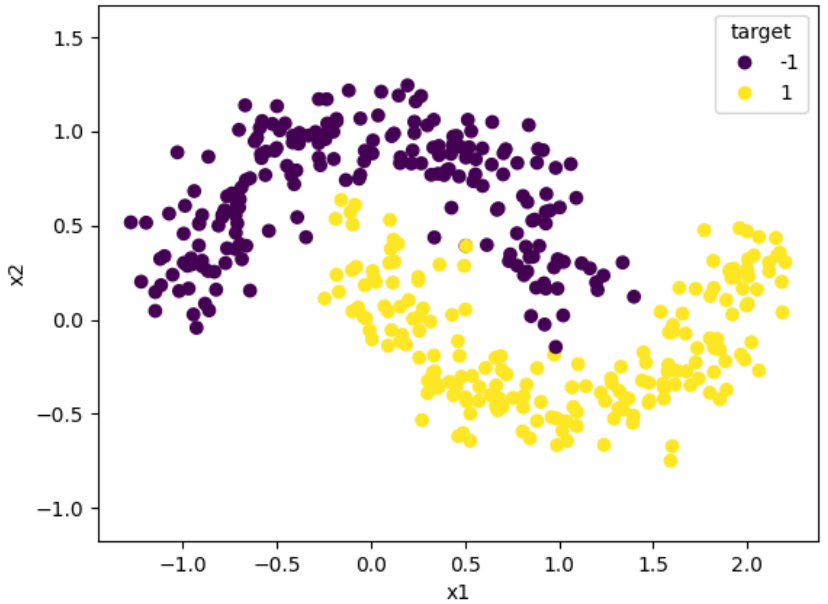
- To find the best possible line, or decision boundary, that separates the data points of different data classes.



https://hackmd.io/@kk6333/S1e7M5RGs

Dataset

- Binary Classification in 2D space
- Features
  - x1
  - x2
- Target
  - target (class label: -1, 1)



Support Vector Machine

- Requirements:
  - Implement the *gram_matrix* function to compute the Gram matrix of the given data with an argument **kernel_function** to specify which kernel function to use.
  - Implement the *linear_kernel* function to compute the value of the linear kernel between two vectors.
  - Implement the *polynomial_kernel* function to compute the value of the polynomial kernel between two vectors with an argument **degree**.
  - Implement the *rbf_kernel* function to compute the value of the rbf kernel between two vectors with an argument **gamma**.

Support Vector Machine

- Tips
  - Your functions will be used in the SVM classifier from scikit-learn like the code below.

```
svc = SVC(kernel='precomputed')
svc.fit(gram_matrix(X_train, X_train, your_kernel), y_train)
y_pred = svc.predict(gram_matrix(X_test, X_train, your_kernel))
```

  - For hyperparameter tuning, you can use any third party library's algorithm to automatically find the best hyperparameter. In your submission, just give the best hyperparameter you used and do not import any additional libraries/packages.

Support Vector Machine

- Criteria:
  - (10%) Show the accuracy score of the testing data using linear_kernel. Your accuracy score should be higher than 0.8.
  - (20%) Tune the hyperparameters of the polynomial_kernel. Show the accuracy score of the testing data using polynomial_kernel and the hyperparameters you used.
  - (20%) Tune the hyperparameters of the rbf_kernel. Show the accuracy score of the testing data using rbf_kernel and the hyperparameters you used.

Points	Testing Accuracy
20 points	**0.98 <= acc**
15 points	**0.90 <= acc < 98**
10 points	**0.85 <= acc < 0.90**
5 points	**0.8 <= acc < 0.85**
0 points	**acc < 0.8**

Code Output

- Do not modify the main function architecture.
- Your code output will look like this:

```
Accuracy of using linear kernel (C = 7):  0.83
Accuracy of using polynomial kernel (C = 1, degree = 3):  0.98
Accuracy of using rbf kernel (C = 10, gamma = 0.1):  0.99
```

Report

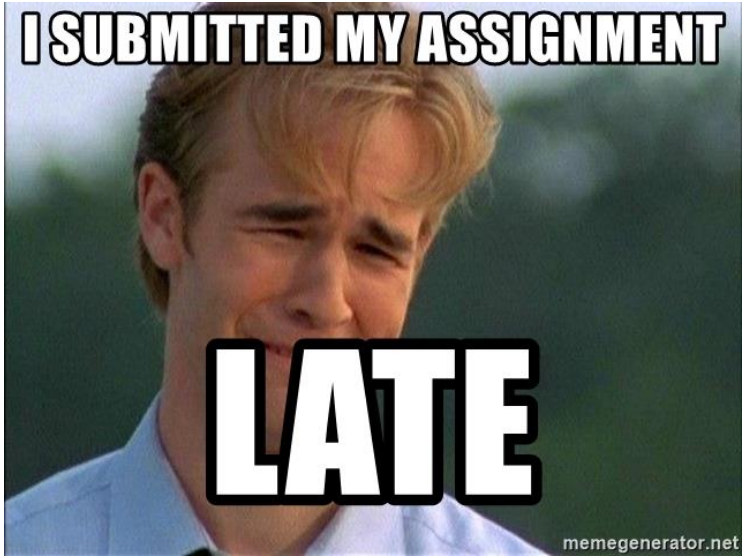
- Please follow the same report template format just like HW1.
- Link

Submission

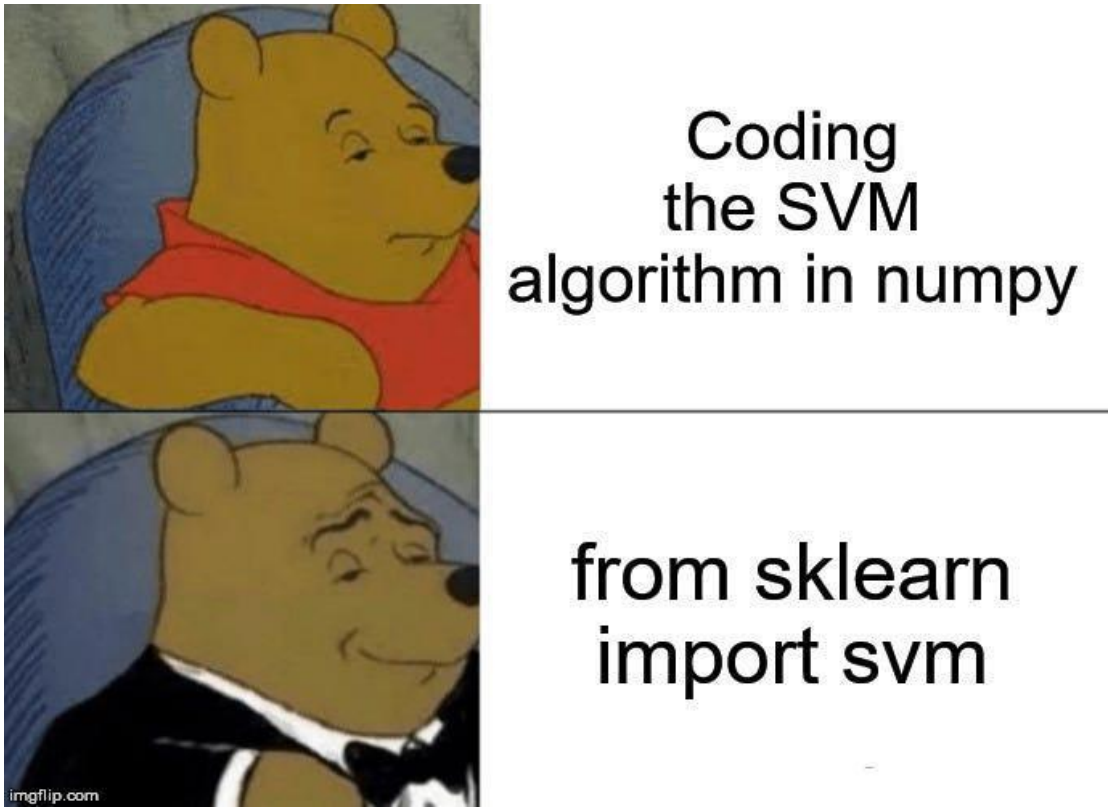
- Compress your code and report into a **.zip file** and submit it on E3.
- <STUDENT ID>_HW4.zip
  - <STUDENT ID>_HW4.py
  - <STUDENT ID>_HW4.pdf (do not submit .doc, .docx or others format)

Late policy

- We will deduct a late penalty of 20 points per additional late day.
- For example, If you get 90 points but delay for two days, your will get only 50 points!



Have Fun



https://towardsdatascience.com/the-complete-guide-to-support-vector-machine-svm-f1a820d8af0b