

Домашнее задание №7: «Ядра SVM»

Дедлайн 1 (20 баллов): 16 апреля, 23:59

Дедлайн 2 (10 баллов): 23 апреля, 23:59

Домашнее задание нужно написать на Python и сдать в виде одного файла. Правило именования файла: `name_surname_7.[py | ipnb]`. Например, если вас зовут Иван Петров, то имя файла должно быть: `ivan_petrov_7.py` или `ivan_petrov_7.ipnb`.

1 Реализуйте Линейный SVM через решение прямой задачи QP для SVM. Рекомендуется воспользоваться пакетом `cvxopt`¹. Также обратите внимание на разреженное представление матрицы `cvxopt.spmatrix`, qр-солвер работает быстрее с разреженными матрицами. Функция `solvers.qp()` решает задачу следующего вида:

$$\begin{cases} \frac{1}{2}x^T P x + q^T x \rightarrow \min_x \\ Gx \leq h \\ Ax = b \end{cases}$$

Необходимо решить следующую задачу оптимизации:

$$\begin{cases} \frac{1}{2}w^T w + C \sum_i \xi_i \rightarrow \min_{w, \xi} \\ \xi_i \geq 0 \\ y_i(w^T x_i + w_0) \geq 1 - \xi_i \quad \forall i = 1 \dots l \end{cases}$$

Сформулируем ее в виде задачи для QP-солвера:

$$x = (w, w_0, \xi)$$

$$P = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} \quad q = \begin{bmatrix} 0 \\ C \cdot 1 \end{bmatrix}$$
$$G = \begin{bmatrix} 0 & 0 & -I \\ -y \odot X & -y & -I \end{bmatrix} \quad h = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

Объект x_i является опорным, если в оптимальной точке для задачи линейного SVM неравенство отступов переходит в равенство:

$$y_i(w^T x_i + w_0) = 1 - \xi_i$$

¹<http://cvxopt.org/userguide/index.html>

Структура класса приведена ниже:

```
class LinearSVM:
    def __init__(self, C):
        self.C = C
        ...

    def fit(self, X, y):
        ...

    def decision_function(self, X):
        ...

    def predict(self, X):
        sign(self.decision_function(X))
```

2 Реализуйте Ядровой SVM через решение двойственной задачи QP для SVM.

Необходимо решить следующую задачу оптимизации:

$$\begin{cases} \sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j K(x_i, x_j) \rightarrow \max_{\alpha} \\ 0 \leq \alpha_i \leq C, \quad \forall i = 1, \dots, l \\ \sum_i \alpha_i y_i = 0 \end{cases}$$

Сформулируем ее в виде задачи для QP-солвера:

$$\begin{aligned} P &= [y_i y_j K(x_i, x_j)] & q &= [-1] \\ G &= \begin{bmatrix} I \\ -I \end{bmatrix} & h &= \begin{bmatrix} C \cdot 1 \\ 0 \end{bmatrix} \\ A &= y^T & b &= 0 \end{aligned}$$

Объект x_i является опорным, если $\alpha_i > 0$.

Решение принимается по следующему правилу:

$$a(x) = \text{sign} \left(\sum_i \alpha_i y_i K(x, x_i) + w_0 \right).$$

Для предсказания необходимо оценить значение w_0 . Известно, что для любого опорного объекта, который классифицируется безошибочно, верно следующее:

$$y_i = \sum_j \alpha_j y_j K(x_i, x_j) + w_0,$$

значит для любого такого объекта:

$$w_0 = y_i - \sum_j \alpha_j y_j K(x_i, x_j)$$

В случае наличия ошибок классификации обучающей выборки предлагается усреднять значение w_0 по всем опорным векторам:

$$w_0 = \frac{1}{N_{SV}} \sum_{n \in SV} \left(y_i - \sum_j \alpha_j y_j K(x_i, x_j) \right).$$

Интуиция здесь такова, что суммарные ошибки в положительную сторону примерно равны суммарным ошибкам в отрицательную сторону.

Структура класса аналогична:

```
class KernelSVM:
    def __init__(self, C, kernel=None, sigma=1.0, degree=2):
        self.C = C
        ...

    def fit(self, X, y):
        ...

    def decision_function(self, X):
        ...

    def predict(self, X):
        ...
```

Параметр конструктора `degree` понадобится при использовании полиномиального ядра, игнорируется другими ядрами. Параметр `sigma` используется для гауссовского ядра. Параметры конструктора соответствуют параметрам стандартного SVM в Scikit-learn ².

3 Реализуйте полиномиальное и гауссовское ядра. Функция должна принимать на вход две матрицы объектов `X` и возвращать матрицу `K`.

Например, линейное ядро выглядит следующим образом:

```
import numpy as np

def linear_kernel(X1, X2):
    return np.dot(X1, X2)
```

4 Воспользуйтесь визуализатором разделяющей поверхности, приведенным по ссылке³ или напишите свой с соответствующей сигнатурой.

```
def visualize(clf, X, y):
    ...
```

5 Протестируйте следующие алгоритмы на случайных выборках с признаками длины 2 и визуализируйте на плоскости получающиеся решающие правила.

1. Линейная разделяющая гиперплоскость обученная линейным SVM и обученная ядровым SVM с линейным ядром
2. Квадратичная разделяющая гиперплоскость (ядровой SVM с полиномиальным ядром степени 2)
3. Radial basis function SVM (ядровой SVM с гауссовским ядром)

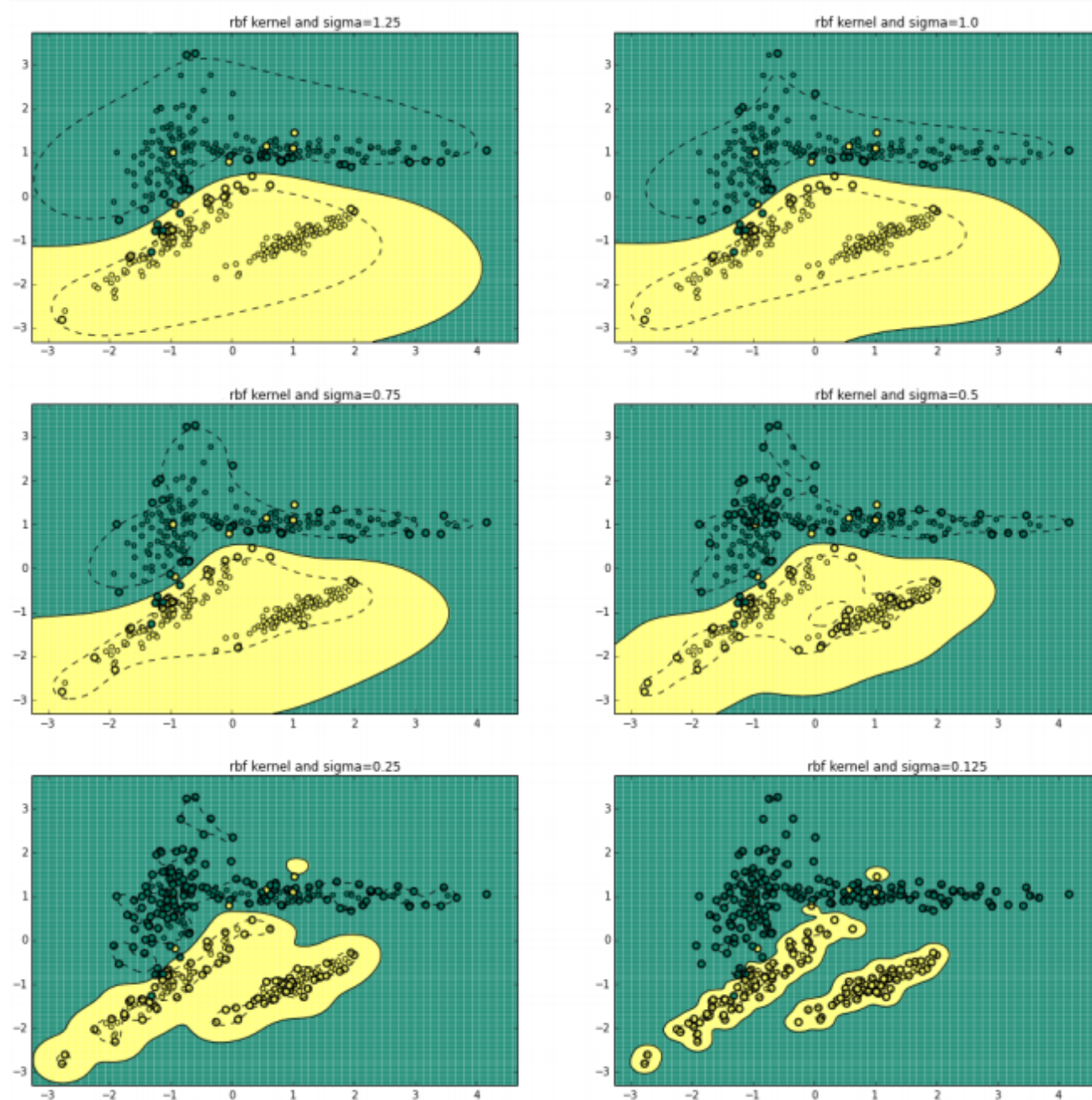
²<http://scikit-learn.org/stable/modules/svm.html>

³<https://gist.github.com/ktisha/ae995b7b553db26316f9>

Для генерации случайных наборов данных можно воспользоваться SciKit-Learn ⁴

6 Поварьируйте параметры классификатора. Что можно сказать про влияние параметров на классификацию?

Пример полученной визуализации для RBF ядра с разными значениями



⁴<http://scikit-learn.org/stable/datasets/#sample-generators>