Homework Assignment

1. Using Python and Numpy, write a class named SVMClassifier, which implements the SVM algorithm having slack variables and kernels such as Polynomial, Gaussian, and Sigmoid (using cvxopt package to solve the quadratic programing problem for Lagrange multipliers).

2. Consider the dataset of letter recognition (named letter-recognition.data). The dataset has 20,000 samples, for which the first column indicates the class (A~Z, totally 26 classes), and the rest columns indicate 16 features as described in the following table. For this dataset, use SVM to do a binary classification for letter ‘C’ or non-‘C’ class.

Homework Assignment

列号	列名	含义	特征/类标记	可取值
1	lettr	A~Z 的字母	类标记	A~Z
2	x-box	horizontal position of box	特征	整数
3	y-box	vertical position of box	特征	整数
4	width	width of box	特征	整数
5	high	height of box	特征	整数
6	onpix	total # on pixels	特征	整数
7	x-bar	mean x of on pixels in box	特征	整数
8	y-bar	mean y of on pixels in box	特征	整数
9	x2bar	mean x variance	特征	整数
10	y2bar	mean y variance	特征	整数
11	xybar	mean x y correlation	特征	整数
12	x2ybr	mean of x * x * y	特征	整数
13	xy2br	mean of x * y * y	特征	整数
14	x-ege	mean edge count left to right	特征	整数
15	xegvy	correlation of x-ege with y	特征	整数
16	y-ege	mean edge count bottom to top	特征	整数
17	yegvx	correlation of y-ege with x	特征	整数