HW5 Tao Ni

2019年4月27日 星期六 上午11:12

1. $L_p = \frac{1}{2} w^T S w - \nu\rho + \sum_t C^t \xi^t$, subj to $\begin{cases} r^t(w^T x^t + w_0) \geq \rho - \xi^t \\ \xi^t \geq 0 \\ \rho \geq 0 \end{cases}$

$\min\limits_{w, \xi, \rho} L_p$

Adding constraint terms to L_p. It becomes:

$$L_p = \frac{1}{2} w^T S w - \nu\rho + \sum_t C^t \xi^t - \sum_t \alpha^t [r^t(w^T x^t + w_0) - \rho + \xi^t] - \sum_t u^t \xi^t - \eta\rho \qquad 1$$

Objective: $\min\limits_{w, \xi, \rho} L_p$

$$\begin{cases} \frac{\partial L_p}{\partial w} = Sw - \sum_t \alpha^t r^t x^t = 0 \Rightarrow Sw = \sum_t \alpha^t r^t x^t \\ \frac{\partial L_p}{\partial w_0} = -\sum_t \alpha^t r^t = 0 \Rightarrow \sum_t \alpha^t r^t = 0 \\ \frac{\partial L_p}{\partial \xi^t} = C^t - \alpha^t - u^t = 0 \Rightarrow \text{since } u^t \geq 0 \quad \therefore \quad \alpha^t \in [0, C^t] \\ \frac{\partial L_p}{\partial \rho} = -\nu + \sum_t \alpha^t - \eta = 0 \Rightarrow \sum_t \alpha^t = \eta + \nu \end{cases}$$

Plug these into 1

$$L_p = \frac{1}{2} w^T \sum_t \alpha^t r^t x^t \cancel{- (\sum_t \alpha^t - \eta)\rho} \cancel{+ \sum_t C^t \xi^t}$$

$$- \sum_t \alpha^t [r^t (w^T x^t \cancel{+ w_0}) \cancel{-\rho} \cancel{+ \xi^t}] \cancel{- \sum_t u^t \xi^t} \cancel{- \eta\rho}$$

$$= \frac{1}{2} w^T \sum_t \alpha^t r^t x^t - \sum_t \alpha^t r^t w^T x^t = -\frac{1}{2} w^T \sum_t \alpha^t r^t x^t \qquad 2$$

Since is positive definite, and $w^T S^T = (\sum_t \alpha^t r^t x^t)^T$

$$2 = -\frac{1}{2} w^T (S^{-1} S)^T \sum_t \alpha^t r^t x^t$$

$$= -\frac{1}{2} (\sum_t \alpha^t r^t x^t)^T \sum_t \alpha^t r^t x^t \cdot (S^{-1})^T$$

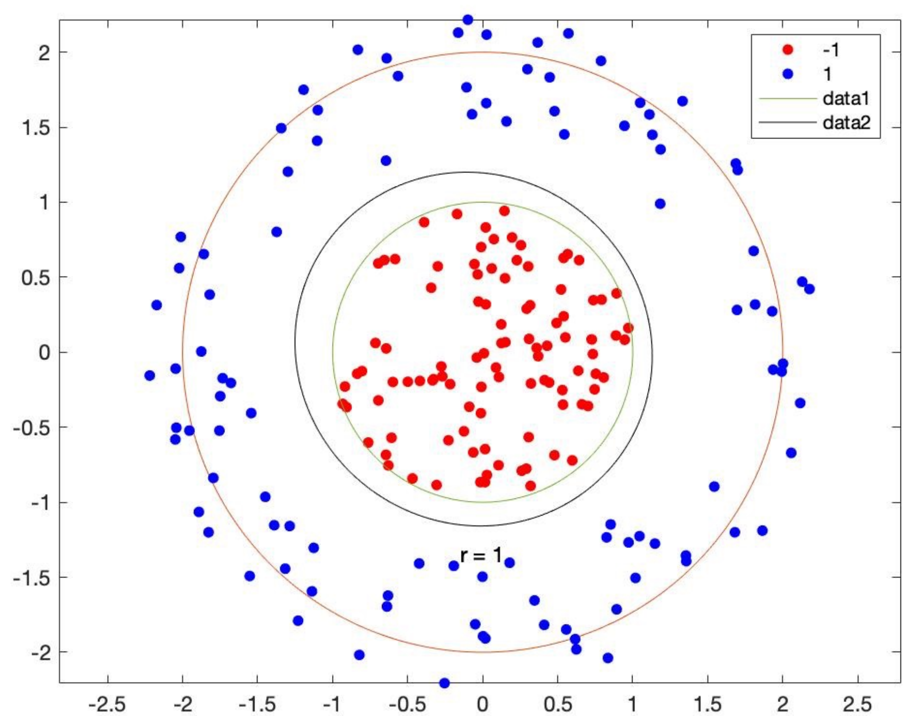
$$= -\frac{1}{2} \sum_t \sum_k \alpha^t \alpha^k r^t r^k (x^k)^T (S^{-1})^T x^t, \text{ subj to } \begin{cases} \sum_t \alpha^t r^t = 0 \\ \alpha^t \in [0, C^t] \\ \sum_t \alpha^t = n + \nu \end{cases}$$

Q2.a

\>> hw5_Q2

"error rate:" "0"

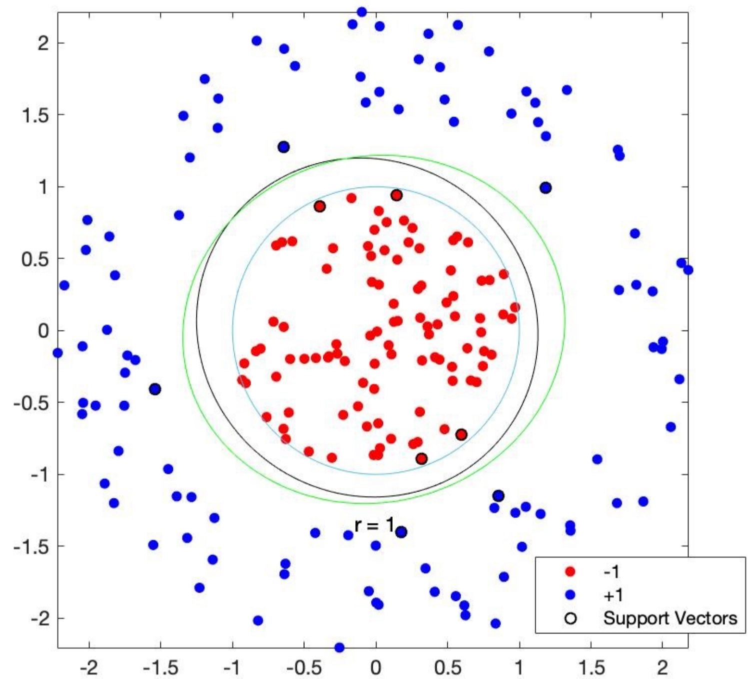
The decision boundary of kernel perceptron is the black contour.



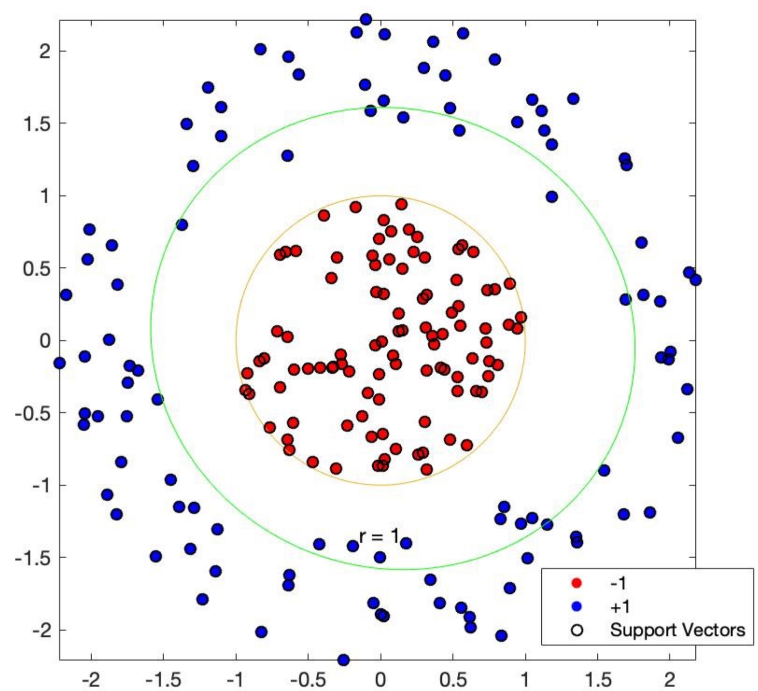
r = 2

Q2.b

The decision boundary of kernel perceptron is the black contour while that of SVM is the green one. We can see from the plot that the decision boundary of SVM is wider than that of my kernel perceptron implementation.



When I set the BoxConstraint to 0.001, I get a plot like this:



We can see from the plot that when I decrease the BoxConstraint parameter, the decision boundary tends to become wider and more and more points are assigned to be the support vectors.

Q2.c

"Train error rate for optdigits49 data"	"0.01182"
"Test error rate for optdigits49 data"	"0.024648"
"Train error rate for optdigits79 data"	"0.013002"
"Test error rate for optdigits79 data"	"0.014184"