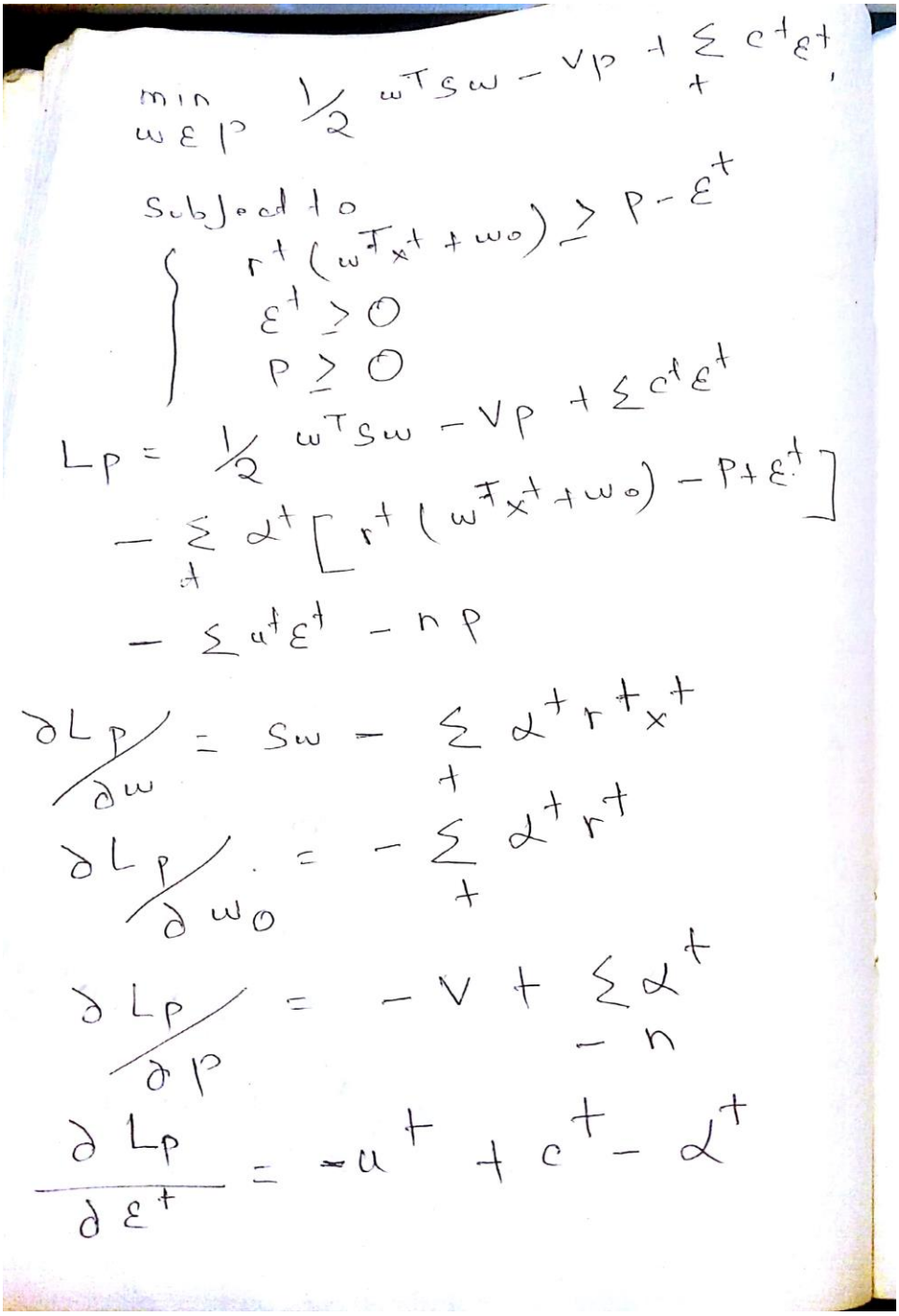


$$\min_{w \in \rho} \frac{1}{2} w^T S w - V\rho + \sum_t c^t \varepsilon^t$$

Subject to

$$\begin{cases} r^t (w^T x^t + w_0) \geq \rho - \varepsilon^t \\ \varepsilon^t \geq 0 \\ \rho \geq 0 \end{cases}$$

$$L_p = \frac{1}{2} w^T S w - V\rho + \sum_t c^t \varepsilon^t - \sum_t \alpha^t \left[r^t (w^T x^t + w_0) - \rho + \varepsilon^t \right] - \sum_t \mu^t \varepsilon^t - n\rho$$

$$\partial L_p / \partial w = Sw - \sum_t \alpha^t r^t x^t$$

$$\partial L_p / \partial w_0 = - \sum_t \alpha^t r^t$$

$$\partial L_p / \partial \rho = -V + \sum_t \alpha^t - n$$

$$\frac{\partial L_p}{\partial \varepsilon^t} = -\mu^t + c^t - \alpha^t$$

$$L_p = \frac{1}{2} w^T S w - \nu \rho + \sum c^t \varepsilon^t$$
$$- \sum \alpha^t \left[r^t (w^T x^t + w_0) - \rho + \varepsilon^t \right]$$
$$- \sum \mu^t \varepsilon^t - \eta \rho$$

$$= \frac{1}{2} w^t \sum_t \alpha^t r^t x^t - \sum \alpha^t \rho + \eta \rho$$
$$\sum c^t \varepsilon^t - \sum \alpha^t \left[r^t (w^T x^t + \overset{0}{\hat{w}_0}) - \rho + \varepsilon^t \right]$$
$$- \sum \mu^t \varepsilon^t - \eta \rho$$

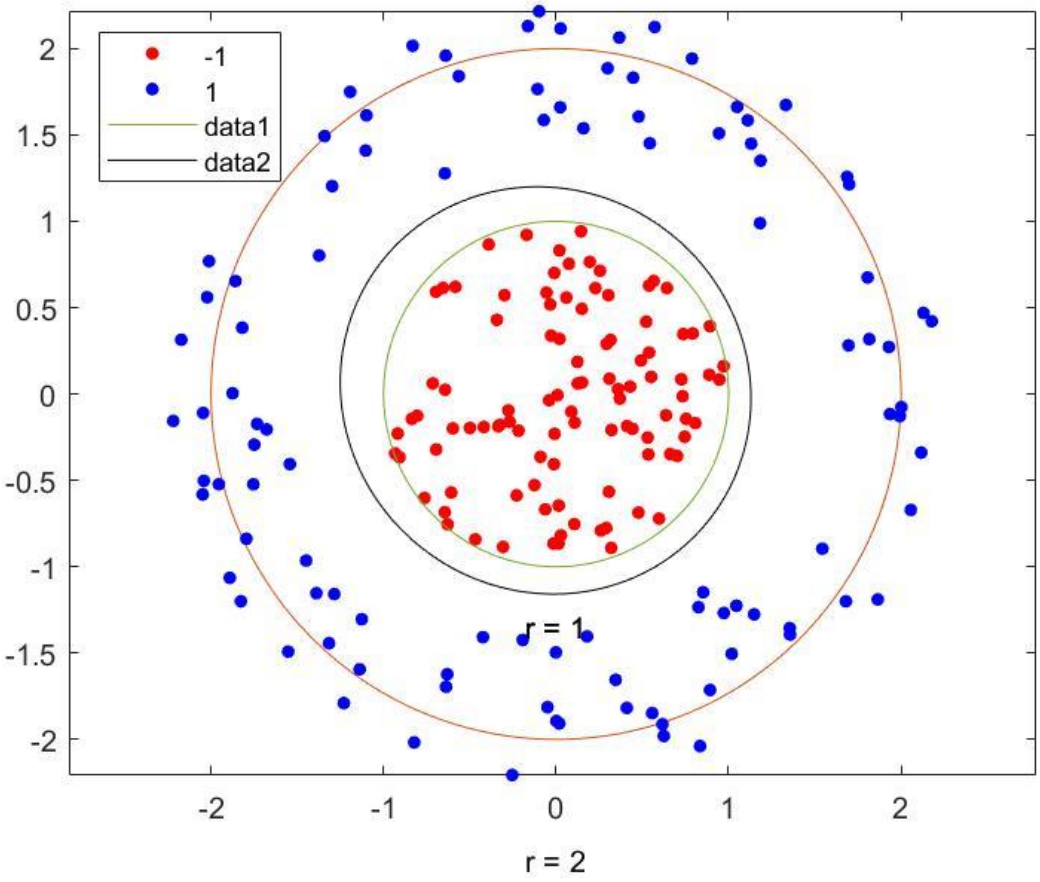
$$= \frac{1}{2} w^t \sum_t \alpha^t r^t x^t - \sum \alpha^t r^t w^t x^t$$

$$= -\frac{1}{2} \sum w^T \alpha^t r^t x^t$$
$$= -\frac{1}{2} \sum w^T S^T S^{-1} x^t \alpha^t r^t$$
$$= -\frac{1}{2} \sum \alpha^t r^t \left[\sum_k \alpha^k r^k x^k \right]^T S^{-1} x^t$$
$$= -\frac{1}{2} \sum_t \sum_k \alpha^t r^t \alpha^k r^k (x^k)^T S^{-1} x^t$$
$$= -\frac{1}{2} \sum_t \sum_k \alpha^t \alpha^k r^t r^k (x^k)^T S^{-1} x^t$$

subj $\sum \alpha^t r^t = 0$, $0 \le \alpha^t \le c^t$, $\sum \alpha^t > \nu$

2 a "the error rate for data3 data"   "0"

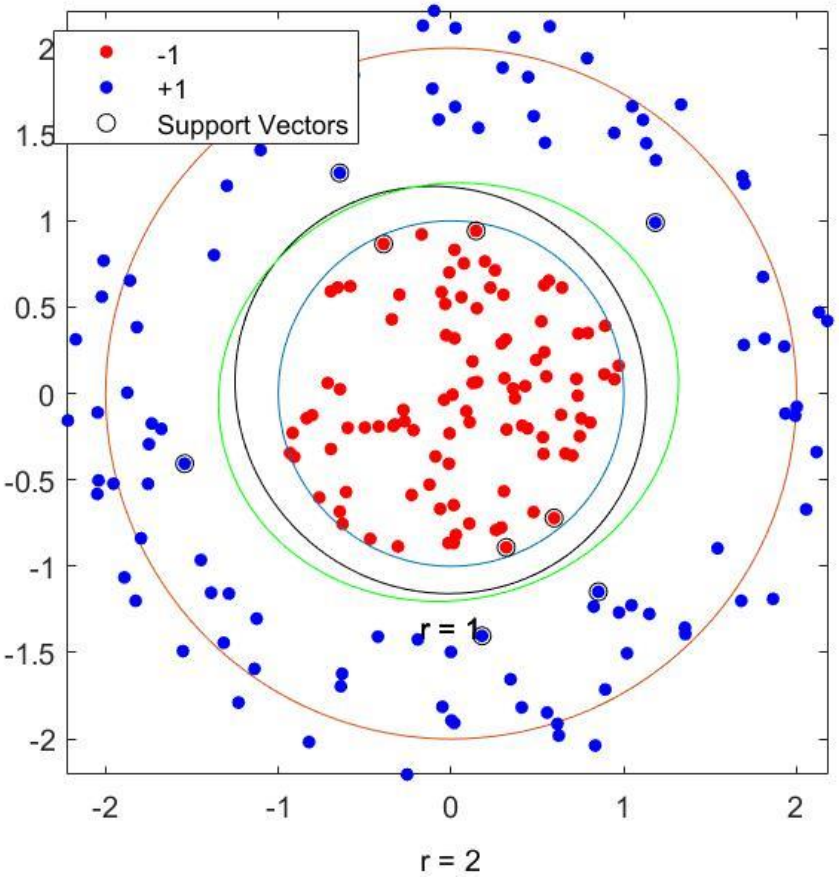
The black circle is formed by the kernel perceptron algorithm



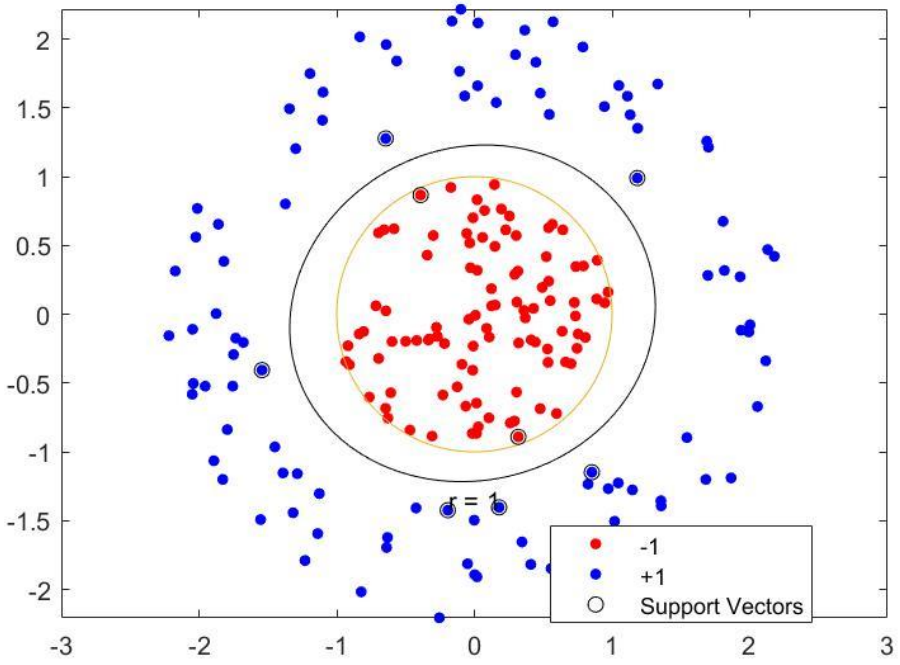
2b

Svm with previous plot(green svm black perceptron)

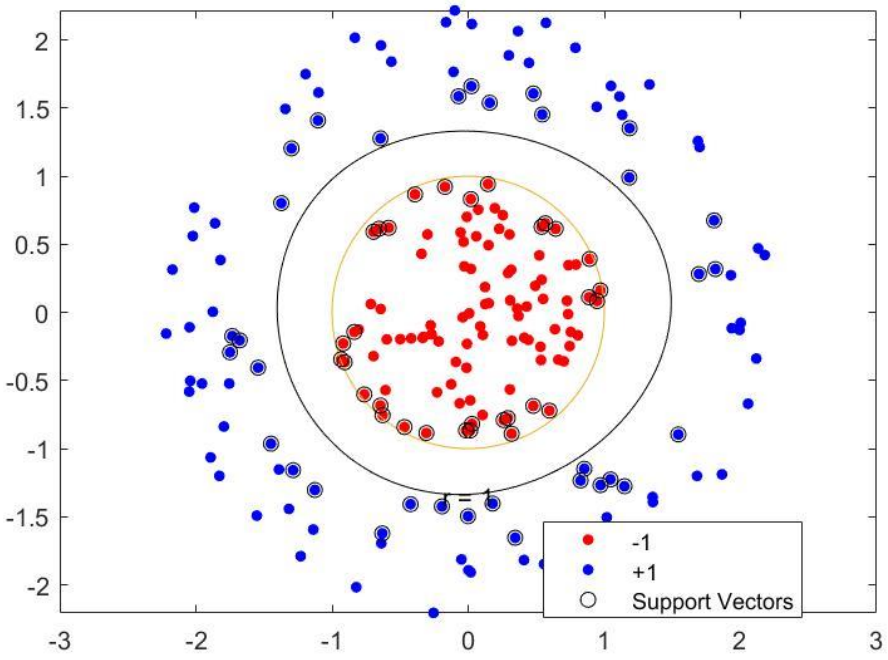
The green boundary of the following curve is svm and the black boundary is perceptron curve.we can see that the green boundary is wider compared to the perceptron boundary.



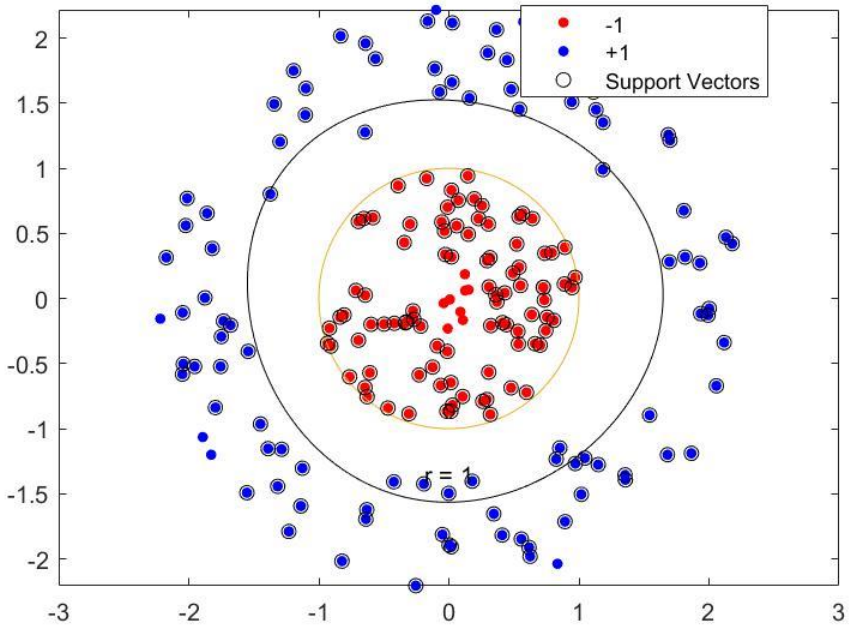
Svm data with box constraint 1 (black curve)



SVM with block constraint 0.001(black curve)



SVM with block constraint 0.001(black curve)



The box constraint means the value of C in object function 1/2*<w,w>+C *Σi si , where C is penalty we choose for misclassification for data. If box constraint or C is large, it means higher cost of the misclassified points, leading to a more strict separation of the data.

2c
Error rates

"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"