

Question 1:

Given: $\min_{w, \xi, \rho} \frac{1}{2} w^T S w - v \rho + \sum_t C^t \xi^t,$

$$\text{subject to } \begin{cases} r^t(w^T x^t + w_0) \geq \rho - \xi^t \\ \xi^t \leq 0 \\ \rho \geq 0 \end{cases}$$

S is positive definite

$$C^t > 0 \forall t$$

$$L_\rho = \frac{1}{2} w^T S w - v \rho + \sum_t C^t \xi^t - \sum_t \alpha^t [r^t(w^T x^t + w_0) - \rho + \xi^t] - \sum_t \mu^t \xi^t - n \rho$$

$$\frac{\partial L_\rho}{\partial w} = S w - \sum_t \alpha^t r^t x^t$$

$$\frac{\partial L_\rho}{\partial w_0} = -\sum_t \alpha^t r^t$$

$$\frac{\partial L_\rho}{\partial \rho} = -v + \sum_t \alpha^t - n$$

$$\frac{\partial L_\rho}{\partial \xi^t} = -\mu^t + C^t - \alpha^t$$

$$\begin{aligned} L_\rho &= \frac{1}{2} w^T \sum_t \alpha^t r^t x^t - \sum_t \alpha^t [r^t(w^T x^t)] \\ &= \frac{1}{2} w^T \sum_t \alpha^t r^t x^t - \sum_t \alpha^t r^t w^T x^t \\ &= -\frac{1}{2} \sum_t w^T \alpha^t r^t x^t \\ &= -\frac{1}{2} \sum_t w^T S^T S^{-1} \alpha^t r^t x^t \\ &= -\frac{1}{2} \sum_t \alpha^t r^t (\sum_k \alpha^k r^k x^k)^T S^{-1} x^t \\ &= -\frac{1}{2} \sum_t \sum_k \alpha^t r^t \alpha^k r^k (x^k)^T S^{-1} x^t \end{aligned}$$

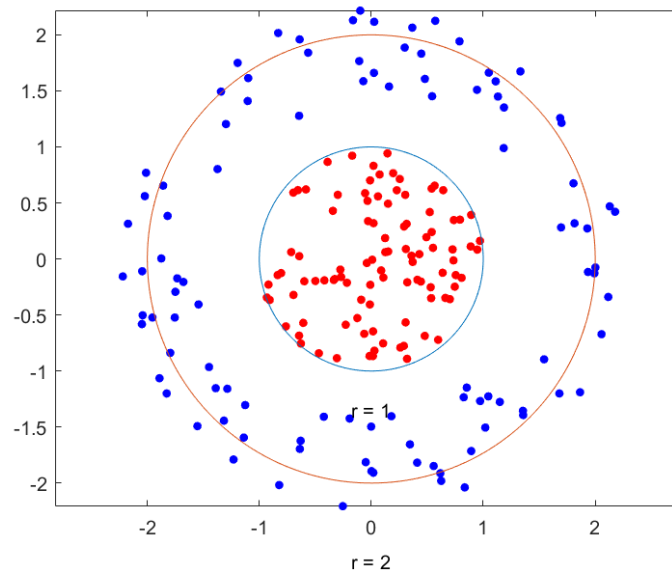
$$= -\frac{1}{2} \sum_t \sum_k \alpha^t \alpha^k r^t r^k (x^k)^T S^{-1} x^t$$

$$\text{subject to } \sum_t \alpha^t r^t = 0, 0 \leq \alpha^t \leq C^t, \sum_t \alpha^t > v, \forall t$$

Question 2a:

Kernel Perceptron

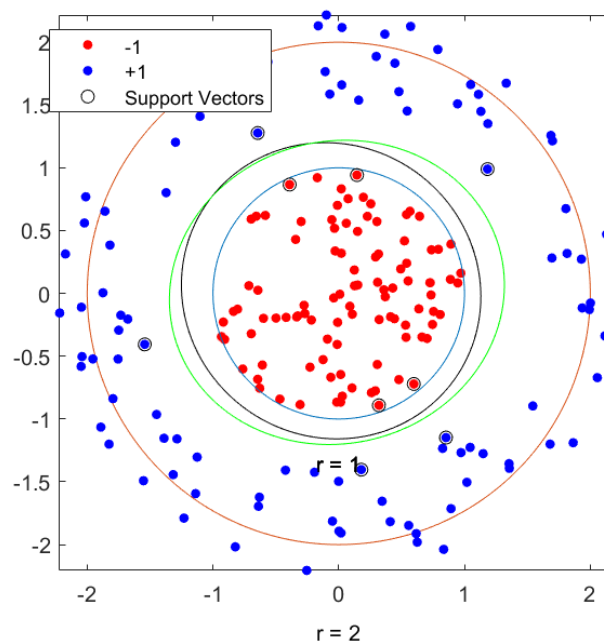
Error Rate: 0



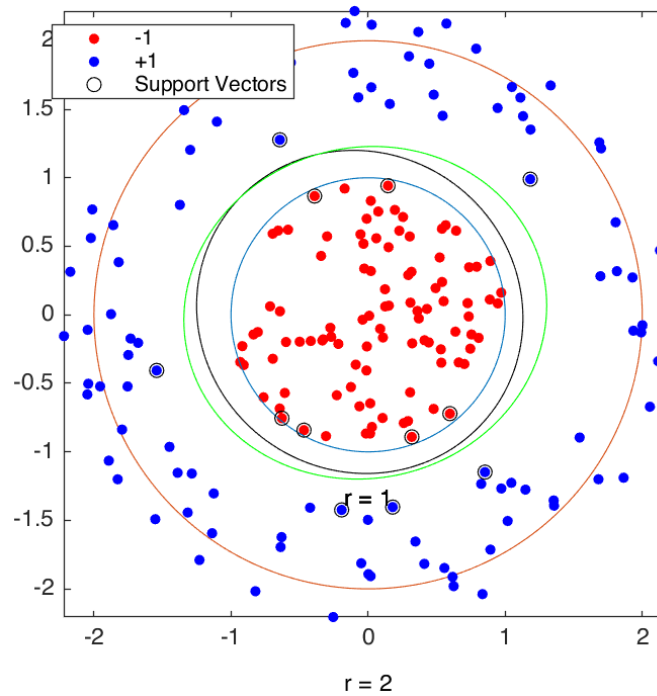
Question 2b:

Kernel Perceptron & SVM

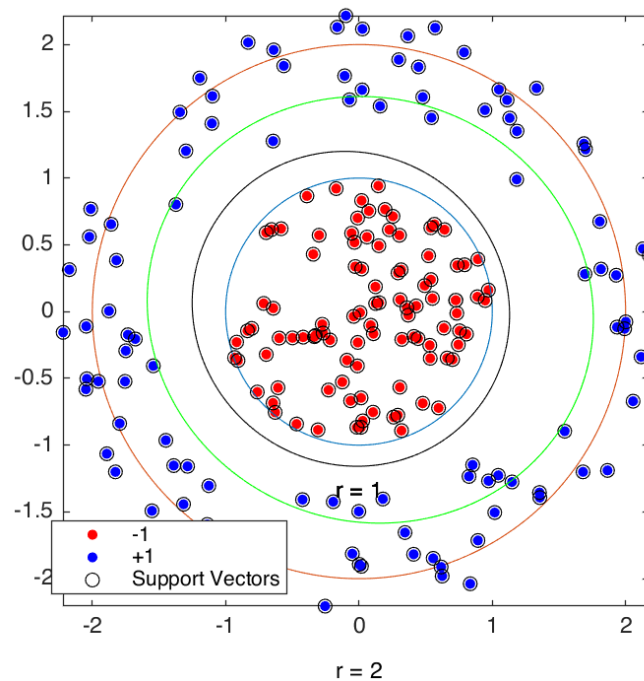
BoxConstraint: 1



BoxConstraint: 0.5



BoxConstraint: 0.001



Increasing the BoxConstraint value is correlated with an increased cost of any misclassified points. This results in a more severe separation of the data.

Question 2c:

optdigits49 Training Error Rate: 0.01182
optdigits49 Testing Error Rate: 0.024648

optdigits79 Training Error Rate: 0.013002
optdigits79 Testing Error Rate: 0.014184