HW 3

(1) Linear kernel since the data is linearly separable.

(2) a). $m/m = 1$

b). $\frac{m/2}{m} = 1/2$.

(3) No because the training error = 0.

Adaboosting cannot really "boost" the performance of this classifier.

(4).

$$G = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 2 & 5 \end{bmatrix}$$

a).

$$\alpha_1 + \alpha_2 + \alpha_3 - \frac{1}{2}\left[\alpha_2^2(1)(1)(1) + \alpha_2\alpha_3(1)(1)(2) + \alpha_3\cdot\alpha_2(1)(1)(2) + \alpha_3^2(1)(1)(5)\right]$$

$$\alpha_1 + \alpha_2 + \alpha_3 - \frac{1}{2}\left[\alpha_2^2 + 4\alpha_2\alpha_3 + 5\alpha_3^2\right]$$

s.t. $\alpha_1, \alpha_2, \alpha_3 \geq 0$, $(\alpha_1)(-1) + (\alpha_2)(1) + \alpha_3(1) = 0$

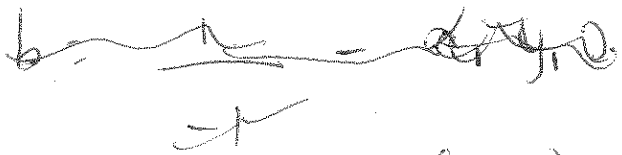
b). $(0, 0)$ & $(0, 1)$ are the support vectors.

c) $z = (-1, 0)$.

$$w^T z = \alpha_1 y_1 x_1^T z + \alpha_2 y_2 x_2^T z$$

$$= 4(-1)(0) + 2(1)(0)$$

$$= 0.$$



Using support vector $(0, 0)$

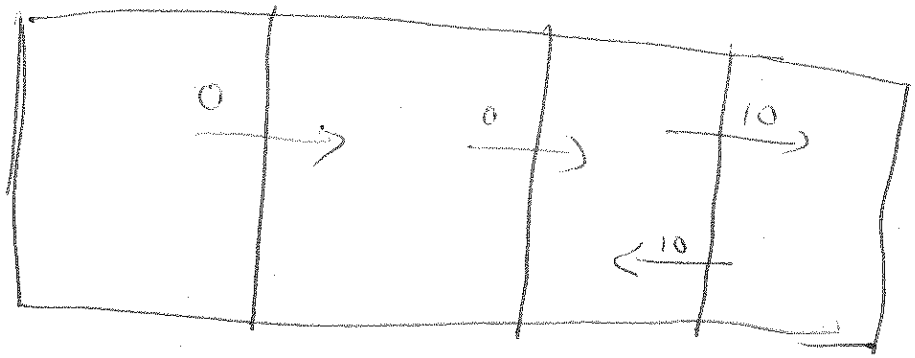
$$b = \frac{1}{-1} - \left[\alpha_1 (-1)(0) + \alpha_2 (1)(0)\right]$$

$$= -1$$

∴ classified as -ve.

c). The data is not linearly separable.
Need to try a different kernel
or Change the "cost" factor in
a soft-margin classifier.

5) i).



ii)

$$Q(A,B) = 0 + \gamma \cdot 0 + \gamma^2 \cdot 10 + \gamma^3 \cdot 10 + \cdots$$

$$= 10\left[\gamma^2 + \gamma^3 + \cdots\right]$$

$$10\gamma^2\left[1 + \gamma + \gamma^2 + \cdots\right]$$

$$10\gamma^2\left[\frac{1}{1-\gamma}\right]$$

$$10\cdot(0.1)^2\left[\frac{1}{1-0.1}\right]$$

$$=$$

$$Q(B,A) = 0 + \gamma \cdot 0 + \gamma^2 \cdot 0 + \gamma^3 \cdot 10 + \gamma^4 \cdot 10 + \cdots$$

$$Q(B,C) = 0 + \gamma \cdot 10 + \gamma^2 \cdot 10 + \gamma^3 \cdot 10 + \cdots$$

$$Q(C,B) = 0 + \gamma \cdot 0 + \gamma^2 \cdot 10 + \gamma^3 \cdot 10 + \cdots$$

6. The idea of increasing the weights of correctly classified examples will not produce the desired result.

Specifically, the main idea of adaboosting is to reduce both bias and variance. Increasing weights of correctly classified points is unlikely to reduce bias since incorrectly classified points are less likely to be classified correctly in a subsequent iteration. The variance of the classifier may reduce since we are averaging across many classifiers.