Problem 1.

1. Proof

The formulation of the maximal margin classifier is given by

$$\arg\max_{\beta, \beta_0} \min_{1 \le i \le n} y_i (\beta^T x_i + \beta_0)$$

To maximize M for every i:

$$y_i (\beta^T x_i + \beta_0) \ge M$$

Scale β to $\|\beta\| = 1$, Therefore

Maximize M subject to $y_i (\beta^T x_i + \beta_0) \ge M$, $\forall i$

2. The Decision function for SVM can be

$$f(x) = \sum_{i=1}^{n} a_i y_i K(x_i, x) + \beta_0$$

For a linear kernel, $K(x_i, x) = x_i^T x$

The a_i are non-zero only for the support vectors, (data points on the margin or misclassified)

Thus, $f(x) = \beta_0 + \sum_{i \in SV} a_i y_i (x_i^T x)$