A linear maximum-margin classifier computes a function

\[ f_{w,w_0}(x) = w^T x + w_0 \]

and assigns a class as \( \text{sgn}(f_{w,w_0}(x)) \) where \( \text{sgn}(x) = 1 \) if \( x \geq 0 \) and \( \text{sgn}(x) = -1 \) otherwise. It is trained using a training sequence \( ((x_1,y_1), \ldots, (x_m,y_m)) \) and the aim in training is to solve the problem

\[
\arg\max_{w,w_0} \left( \min_i \frac{y_i f_{w,w_0}(x_i)}{||w||} \right). \tag{1}
\]

(a) Give a brief explanation of how Equation 1 is derived. You may assume that the distance from \( x' \) to the line \( f_{w,w_0}(x) = 0 \) is \( |f_{w,w_0}(x')|/||w|| \). [2 marks]

(b) Why is Equation 1 not used in practice? Explain how an alternative optimization problem is derived that can form the basis of a practical learning algorithm. You need only derive a statement of the primal optimization problem. [3 marks]

(c) Explain how the linear maximum-margin classifier can be modified to be nonlinear and to allow misclassification of the training examples. Give a derivation of the modified optimization problem needed for training. You need only derive a statement of the primal optimization problem. [4 marks]

(d) As part of the derivation of the full learning algorithm we find that the function \( f \) might be expressible in terms of \( m \) new parameters \( \alpha_i \) as

\[
 f_{\alpha_1,\ldots,\alpha_m,w_0}(x) = \sum_{i=1}^{m} y_i \alpha_i K(x_i,x) + w_0. \tag{2}
\]

Explain the purpose of \( K \) in Equation 2 and explain why its use might be beneficial. [3 marks]

(e) Your boss can not afford to provide you with a solver capable of training your system using the algorithm in Parts (c) and (d). Your boss does however provide you with a solver for linear programs. For a matrix \( A \) and vectors \( b \) and \( c \), this solves problems of the form

Find \( x \) minimizing \( b^T x \) with constraints \( Ax \geq c \) and \( x \geq 0 \).

Suggest a way in which you could use this optimizer to (approximately) train your system. [8 marks]