The standard linear regression model uses a hypothesis
\[ h(x, w, b) = w^T x + b \]
to fit \( m \) examples \( ((x_1, y_1), \ldots, (x_m, y_m)) \) by minimizing the error
\[ E(w, b) = \sum_{i=1}^{m} (y_i - h(x_i, w, b))^2. \]

(a) Derive a gradient descent algorithm for training the linear regression model described. [5 marks]

(b) In the application of interest, you believe that it is desirable to train such that the learned parameters have \( ||w|| \approx 1 \). Suggest a modification to \( E(w, b) \) that facilitates this, and derive the corresponding gradient descent training algorithm. [5 marks]

(c) Describe the components of a general heuristic search problem. [4 marks]

(d) You are faced with a heuristic search problem, but the heuristics you have so far developed are less effective than desired. Suggest two ways in which supervised machine learning might be used to develop a better heuristic, mentioning if necessary any corresponding disadvantages of using the approach. You may assume that a collection of problems to be solved by the heuristic search is available. [6 marks]