**1   Artificial Intelligence (sbh11)**

The standard *linear regression* model uses a hypothesis

$$h(\mathbf{x}, \mathbf{w}, b) = \mathbf{w}^T \mathbf{x} + b$$

to fit $m$ examples $((\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_m, y_m))$ by minimizing the error

$$E(\mathbf{w}, b) = \sum_{i=1}^{m} (y_i - h(\mathbf{x}_i, \mathbf{w}, b))^2 .$$

(*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 $||\mathbf{w}|| \simeq 1$. Suggest a modification to $E(\mathbf{w}, b)$ 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]