You have a labelled data set \( s = ((x_1, y_1), \ldots, (x_m, y_m)) \) with \( x_i \in \mathbb{R}^n \) and \( y_i \in \{+1, -1\} \). The maximum margin classifier computes

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

\[
h_{w_0}(x) = \text{sgn}(f_{w_0}(x))
\]

where \( \text{sgn}(x) = +1 \) if \( x > 0 \) and \( \text{sgn}(x) = -1 \) otherwise.

(a) One approach to training the maximum margin classifier would be to solve the problem

\[
(w, w_0) = \text{argmax} \left[ \min_i \frac{y_i f_{w_0}(x_i)}{||w||} \right].
\]

Explain how this version of the training algorithm is derived, paying particular attention to the meaning of the term \( f_{w_0}(x_i)/||w|| \). [5 marks]

(b) Explain why the training algorithm in Part (a) is not used in practice. [1 mark]

(c) Describe in detail two alternative ways of formulating the training of the maximum margin classifier as a constrained optimization problem. You need not describe an algorithm for solving the constrained optimization, but should explain in each case how a combination of objective function and constraints is obtained from first principles. [7 marks]

(d) Evil Robot has completed a course on some software called VectorDribble, and now considers himself a Data Science Expert. He claims that, as the support vector machine and Gaussian process regressor both involve a function of the form \( K : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} \), they are essentially the same method. Explain, in as much detail as you can, why Evil Robot is mistaken. [7 marks]