

8 Machine Learning and Bayesian Inference (sbh11)

You have a labelled data set $\mathbf{s} = ((\mathbf{x}_1, y_1), \dots, (\mathbf{x}_m, y_m))$ with $\mathbf{x}_i \in \mathbb{R}^n$ and $y_i \in \{+1, -1\}$. The *maximum margin classifier* computes

$$f_{\mathbf{w}, w_0}(\mathbf{x}) = w_0 + \mathbf{w}^T \Phi(\mathbf{x})$$

$$h_{\mathbf{w}, w_0}(\mathbf{x}) = \text{sgn}(f_{\mathbf{w}, w_0}(\mathbf{x}))$$

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

$$(\mathbf{w}, w_0) = \operatorname{argmax} \left[\min_i \frac{y_i f_{\mathbf{w}, w_0}(\mathbf{x}_i)}{\|\mathbf{w}\|} \right].$$

Explain how this version of the training algorithm is derived, paying particular attention to the meaning of the term $f_{\mathbf{w}, w_0}(\mathbf{x}_i)/\|\mathbf{w}\|$. [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 $K : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$, they are essentially the same method. Explain, in as much detail as you can, why Evil Robot is mistaken. [7 marks]