Solution notes

Artificial Intelligence I 2005 – Paper 4 Question 4 (SBH)

This addresses the section of the course devoted to neural networks. Specifically the section which introduces the concept of a kernel without resorting to the full machinery of support vector machines.

(*a*) Set $\eta = 1$ so the algorithm works by adding or subtracting misclassified $\mathbf{x}_i$ to $\mathbf{w}$. Then

$$\mathbf{w} = \sum_{i=1}^{m} \alpha_i y_i \mathbf{x}_i$$

where the α_i are positive and depend on the number of times $\mathbf{x}_i$ was misclassified. The vector α of these values provides an alternative to $\mathbf{w}$. The two lines in the `if` in the algorithm are replaced by $\alpha_i = \alpha_i + 1$ and $w_0 = w_0 + y_i R^2$.

(*b*) Introduce d typically nonlinear basis functions $\phi_i : \mathbf{R}^n \to \mathbf{R}$ and map each $\mathbf{x}_i \in \mathbf{R}^n$ to a new space $\mathbf{R}^d$ then use a perceptron in $\mathbf{R}^d$ so

$$h(\mathbf{x}) = \operatorname{sgn}\left(\sum_{i=1}^{d} w_i \phi_i(\mathbf{x}) + w_0\right)$$

(*c*) A function K such that $K(\mathbf{x}, \mathbf{y}) = \Phi(\mathbf{x})^T \Phi(\mathbf{y})$ where $\Phi(\mathbf{x})^T = (\phi_1(\mathbf{x}), \ldots, \phi_p(\mathbf{x}))$ for basis functions ϕ_i.

(*d*) Writing h for the dual algorithm but no basis functions gives

$$h(\mathbf{x}) = \operatorname{sgn}\left(\sum_{i=1}^{m} \alpha_i y_i (\mathbf{x}_i^T \mathbf{x}) + w_0\right)$$

For the primal algorithm and using basis functions we have

$$h(\mathbf{x}) = \operatorname{sgn}\left(\sum_{i=1}^{d} w_i \phi_i(\mathbf{x}) + w_0\right)$$

For the dual algorithm with basis functions we have

$$h(\mathbf{x}) = \operatorname{sgn}\left(\sum_{i=1}^{m} \alpha_i y_i \Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}) + w_0\right)$$

where the inner product can be replaced by a kernel K. There is thus a trade off involving m and d. As d can be huge there is a major benefit to using the dual algorithm with a kernel provided the kernel makes the computation of the inner product relatively easy.