2005 Paper 4 Question 4

Artificial Intelligence I

A *perceptron* computes the function $h(\mathbf{x}) = \mathrm{sgn}(\mathbf{w}^T\mathbf{x} + w_0)$ where $\mathrm{sgn}(x) = +1$ if $x \geq 0$ and $\mathrm{sgn}(x) = -1$ otherwise. The *primal perceptron algorithm* is as follows:

```
do
{
    for (each example in s)
    {
        if (yi(wᵀxi + w0) ≤ 0)
        {
            w = w + ηyixi
            w0 = w0 + ηyiR²
        }
    }
}
while (mistakes are made in the for loop)
```

where η is a positive real, $R = \max_i \|\mathbf{x}_i\|$ and $\mathbf{w}$ and w_0 are initialised to be the zero vector and 0 respectively, and where

$$\mathbf{s} = ((\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_m, y_m)) \text{ with } y_i \in \{+1, -1\}$$

is a training sequence.

(*a*) Derive the dual form of the perceptron algorithm and state the corresponding alternative representation for $\mathbf{w}$. [6 marks]

(*b*) Explain how the perceptron may be applied to problems that are not linearly separable by introducing *basis functions*. [4 marks]

(*c*) Give a definition of a *kernel*. [3 marks]

(*d*) Explain how the use of a suitable kernel in conjunction with the dual form of the perceptron algorithm can be advantageous compared with the direct use of basis functions and the primal perceptron algorithm. [7 marks]