Prob Set #3

5.

binary classification $y \in \{0, 1\}$

$h_\theta(x) = g(\theta^T x)$, where $g(z) = \begin{cases} 1 & \text{if } z \geq 0 \\ 0 & \text{otherwise} \end{cases}$

<update rule>

$$\theta^{(i+1)} := \theta^{(i)} + \alpha(y^{(i+1)} - h_{\theta^{(i)}}(x^{(i+1)}))x^{(i+1)}$$

(a)

$\phi \in \mathbb{R}^\infty$

Describe how you would apply "kernel trick" to the perceptron to make it work in the high-dimensional feature space ϕ, but without ever explicitly computing $\phi(x)$

m. size of training data

$$f(x) = \sum_{j=1}^{m} \beta_j y^{(j)} K(x^{(j)}, x) + b \qquad K \to \mathbb{R}$$

$$h_{\theta^{(i)}}(x^{(i+1)}) = g(\theta^{(i)T} \phi(x^{(i+1)}))$$

$\theta^{(i)}$: linear combination of $\phi(x^{(1)}), \dots, \phi(x^{(i)})$

i) $$\theta^{(i)} = \sum_{j=1}^{i} \beta_j \phi(x^{(j)})$$

$$\theta^{(0)} = \sum_{j=1}^{0} \beta_j \phi(x^{(j)}) = 0$$

ii) $$h_{\theta^{(i)}}(x^{(i+1)}) = g\left(\left(\sum_{j=1}^{i} \beta_j \phi(x^{(j)})\right)^T \phi(x^{(i+1)})\right)$$

$$= g\left(\sum_{j=1}^{i} \beta_j K(x^{(j)}, x^{(i+1)})\right)$$

iii) $$\theta^{(i+1)} := \theta^{(i)} + \alpha(y^{(i+1)} - h_{\theta^{(i)}}(x^{(i+1)}))\phi(x^{(i+1)})$$

$$= \sum_{j=1}^{i} \beta_j \phi(x^{(j)}) + \alpha\left(y^{(i+1)} - g\left(\sum_{j=1}^{i} \beta_j K(x^{(j)}, x^{(i+1)})\right)\right)\phi(x^{(i+1)})$$