CS691 Quiz 11 Name Solution

1. What is the "kernel trick"? And why is it useful? (3 points)

$K(\vec{x},\vec{z}) = \phi(\vec{x})\cdot\phi(\vec{z})$

The kernel trick allows us to replace dot products between samples w/ the kernel product. We are able to take advantage of nonlinearity without the extra computational complexity

2. What does α replace in the kernelized perceptron algorithm? (2 points)

w

3. I have N training samples with D features. How many dimensions is w? What about α? (2 points)

w = D dim

α = N dim

4. I have 10 training samples, $s_1, \ldots, s_{10}$. $s_i = (x_i, y_i)$, so s_i has a feature vector x_i and a label y_i. If while running the kernelized perceptron algorithm, I get s_3, s_5, and s_{10} wrong in the first epoch, then s_4, and s_{10} wrong in the second epoch, and then nothing wrong in the third epoch, what is α? (3 points)

If you understand the kernelized perceptron, you do not need the following equations. They are provided for you so you can answer the question even if you do not quite understand the algorithm. The first equation is w in terms of α, and the second equation is the update rule for the original perceptron algorithm.

$$w = \sum_n^N \alpha_n \phi(x_n)$$

$$w = w + y_n x_n$$

$\alpha = (0, 0, y_3, y_4, y_5, 0, 0, 0, 0, 2y_{10})$

CS491 Quiz 11 Name ______________________

1. What is the "kernel trick"? And why is it useful? (3 points)

Same as 691

2. I have N training samples with D features. How many dimensions is w? What about α? (2 points)

Same as 691

3. If $w = (0, 0, \ldots, 0)$, what is α? (2 points)

$\alpha = \vec{0}$

4. I have 10 training samples, $s_1, \ldots, s_{10}$. $s_i = (x_i, y_i)$, so s_i has a feature vector x_i and a label y_i. If while running the kernelized perceptron algorithm, I get s_3, s_5, and s_{10} wrong in the first epoch, then s_4, and s_{10} wrong in the second epoch, and then nothing wrong in the third epoch, what is α? (3 points)
If you understand the kernelized perceptron, you do not need the following equations. They are provided for you so you can answer the question even if you do not quite understand the algorithm. The first equation is w in terms of α, and the second equation is the update rule for the original perceptron algorithm.

$$w = \sum_{n}^{N} \alpha_n \phi(x_n)$$

$$w = w + y_n x_n$$

Same as 691