

## Tutorial 6

### Detecting spam mails

One of the fundamental tasks of machine learning is to detect spam e-mails. You are given some words and a label of +1 if it is spam or -1 if it is not. Here **1** indicates the presence of word and **0** the absence of word. Assume the learning rate  $\mathcal{T}$  is  $\frac{1}{2}$ . Find the separating hyperplane using perceptron training algorithm

	area	sexy	your	in	singles	y
a	1	1	0	1	1	+1
b	0	0	1	1	0	-1
c	0	1	1	0	0	+1
d	1	0	0	1	0	-1
e	1	0	1	0	1	+1
f	1	0	1	1	0	-1

### Computing power of perceptrons

Perceptrons can only separate Linearly separable data as discussed in class. Given  $n$  variables we can have  $2^{2^n}$  boolean functions, but not all of these can be represented by a perceptron. For example when  $n=2$  the XOR and XNOR cannot be represented by a perceptron. Given  $n$  boolean variables how many of  $2^{2^n}$  boolean functions can be represented by a perceptron?

### Kernel Perceptron

Recall the proof for convergence of the perceptron update algorithm. Now can this proof be extended to the kernel perceptron? Kernelized perceptron:

- INITIALIZE:  $\alpha = \text{zeros}()$
- REPEAT: for  $\langle x_i, y_i \rangle$ 
  - If  $\text{sign}\left(\sum_j \alpha_j y_j K(x_j, x_j)\right) \neq y_i$
  - then,  $\alpha_j = \alpha_j + 1$
  - endif

## Number of iterations for convergence of perceptron update

Prove the following:

If  $\|w^*\| = 1$  and if there exists  $\theta > 0$  such that for all  $i = 1, \dots, n$ ,  $y_i(w^*)^T \phi(x_i) \geq \theta$  and  $\|\phi(x_i)\|^2 \leq \Gamma^2$  then the perceptron algorithm will make at most  $\frac{\Gamma^2}{\theta^2}$  errors.