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 seperating 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

Solution: This is a programming exercise. Pls share and discussions solutions to this problem on tutorspace.

Computing power of perceptrons

Perceptrons can only seperate Linearly seperable 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?

Solution: Again, suggest you discuss solution to this on Tutorspace before I post the solution.

Kernel Perceptron

Recall the proof for convergence of the perceptron update algorithm. Now can this proof be extended to the kernel perceptron? Kernelized perceptron[1]:

$$f(x) = sign\left(\sum_i \alpha_i^* y_i K(x, x_i) + b^*\right)$$

[1] In the original tutorial problem, b was missing. Re-introducing b helps state the equivalence of kernel perceptron to regular perceptron more easily.

- INITIALIZE: α=zeroes()
- REPEAT: for $< x_i, y_i >$
  - If $sign\left(\sum_j \alpha_j y_j K(x_j, x_j) + b\right) \neq y_i$
  - then, $\alpha_j = \alpha_j + 1$
  - endif

Solution: Yes, in fact kernel perceptron can be derived from the perceptron update rule as follows:

$f(x) = sign\left((w^*)^T \phi(x)\right) = sign\left(\sum_i \alpha_i^* y_i K(x, x_i) + b^*\right)$

- INITIALIZE: $w = [0, 0, \ldots, 0, 1] \Rightarrow f(x) = sign\left((w)^T \phi(x)\right) = sign\left(\sum_i \alpha_i y_i K(x, x_i) + b\right)$

  with $\alpha_i = 0$ and $b = 1$

  Note: $\phi^T(\widehat{x})\phi(x)\widehat{y} = \widehat{y}K(\widehat{x}, x) + \widehat{y}$
- REPEAT: for each $< \widehat{x}, \widehat{y} >$
  - If $\widehat{y}w^T\phi(\widehat{x}) < 0$

    $\Rightarrow f(\widehat{x}) = sign\left((w)^T \phi(\widehat{x})\right) = sign\left(\sum_i \alpha_i y_i K(\widehat{x}, x_i) + b\right) \neq \widehat{y}$
  - then, $w' = w + \Phi(\widehat{x}).\widehat{y}$

    $\Rightarrow f(x) = sign\left((w')^T \phi(x)\right) = sign\left(\sum_i (\alpha_i y_i K(x, x_i) + \phi^T(\widehat{x})\phi(x)\widehat{y}) + b\right)$

    $= sign\left(\sum_i \alpha'_i y_i K(x, x_i) + b'\right)$ where $\alpha'_i = \alpha_i$ for all i except that $\alpha'_{\widehat{x}} = \alpha_{\widehat{x}} + 1$ and $b' = b + \widehat{y}$
  - endif

Thus, $f(x) = sign\left((w^*)^T \phi(x)\right) = sign\left(\sum_i^* \alpha_i y_i K(x, x_i)\right)$

Number of iterations for convergence of perceptron update

Prove the following:

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

Solution: `http://www.cs.columbia.edu/~mcollins/courses/6998-2012/notes/perc.converge.pdf`