

# Homework 1

Due: 16th June 2014

**Problem 1** Let  $X$  be a random variable which denotes if a patient has AIDS and let  $T_1$  and  $T_2$  be the outcomes of two clinical tests with the following error profile:

| $p(t_1 x)$          | $x = \text{HIV-}$ | $x = \text{HIV+}$ | $p(t_2 x)$          | $x = \text{HIV-}$ | $x = \text{HIV+}$ |
|---------------------|-------------------|-------------------|---------------------|-------------------|-------------------|
| $t_1 = \text{HIV-}$ | 0.91              | 0.07              | $t_2 = \text{HIV-}$ | 0.98              | 0.01              |
| $t_1 = \text{HIV+}$ | 0.09              | 0.93              | $t_2 = \text{HIV+}$ | 0.02              | 0.99              |

Compute

1.  $p(x = +|t_1 = +, t_2 = +)$
2.  $p(x = +|t_1 = +, t_2 = -)$
3.  $p(x = +|t_1 = -, t_2 = +)$
4.  $p(x = +|t_1 = -, t_2 = -)$ .

You may assume that the prior probability  $p(x = +)$  is 0.005.

**Problem 2** Download any *two* binary classification datasets from the following website <http://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/binary.html>. One of them should have ordinal attributes and the second one should have continuous attributes. Train a Naive Bayes classifier using the two datasets and test the performance in terms of accuracy. I am looking for a brief report (1 page max) on how you approached the problem, what results you obtained, what practical difficulties you faced, and how you overcame these difficulties.