T.61.5140 Machine Learning: Advanced Probablistic Methods
Hollmén, Raiko (Spring 2008)
Problem session, 29th of February, 2008
http://www.cis.hut.fi/Opinnot/T-61.5140/

1. Given a Naïve Bayes model with four binary variables C, X_1, X_2, X_3, that is $P(C, X_1, X_2, X_3) = P(C)P(X_1 \mid C)P(X_2 \mid C)P(X_3 \mid C)$ and a dataset with five samples $t = 1 \ldots 5$ (see table below), write the likelihood function $P(C, X_1, X_2, X_3 \mid \boldsymbol{\theta})$ of the model parameters $\boldsymbol{\theta}$ (the values in the conditional probability tables). Find $P(C)$ and $P(X_1 \mid C = 1)$ that maximize the likelihood (use the notation $\theta_1 = P(C = 1)$ and $\theta_2 = P(X_1 = 1 \mid C = 1)$).

Data:

t	C_t	X_{1t}	X_{2t}	X_{3t}
1	0	1	0	1
2	1	0	1	0
3	0	0	1	0
4	1	0	1	1
5	1	1	1	0

2. Given a Naïve Bayes model with three binary variables defined by the tables below, classify the data set below. Classification is defined as $C^* = \arg\max_C P(C \mid X_1, X_2)$.

P(C)	
C=0	0.7
C=1	0.3

$P(X_1 \mid C)$	C=0	C=1
X_1=0	0.5	0.8
X_1=1	0.5	0.2

$P(X_2 \mid C)$	C=0	C=1
$X_2 = 0$	0.6	0.3
$X_2 = 1$	0.4	0.7

Data:

t	X_{1t}	X_{2t}
1	1	1
2	0	1

3. Run an iteration of the EM-algorithm for the model in Problem 2. The E-step is to compute $P(C \mid X_1, X_2)$ and the M-step is to maximize the expected likelihood over the distribution from the E-step.

4. (a) Run k-means (page 424) until convergence in a one-dimensional problem with five data points (see table below). Use $k = 2$ and initialize with $\mu_1 = 3.5$ and $\mu_2 = 4.8$. (b) Fit a mixture-of-Gaussians (MoG, page 430) to the result. MoG is a model with a cluster label C and a Gaussian distribution for the observation given the cluster label:

$$p(x \mid C = i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left[-\frac{(x-\mu_i)^2}{2\sigma_i^2}\right]. \tag{1}$$

You can fit the Gaussians by computing the mean $\mu = E(x)$ and variance $\sigma^2 = E(x^2) - E(x)^2$ of the data in each cluster. (c) Compute $P(C \mid x = 3)$.

Data:

t	x_t
1	1.0
2	2.0
3	4.0
4	5.0
5	6.0