Lecture Graphical Models
`https://ml01.zrz.tu-berlin.de/wiki/Main/SS09_GraphicalModels`
Machine Learning Group, TU Berlin

Instructors: Dr. Ulf Brefeld, Dr. Marc Toussaint
Tutor: Tobias Lang, `lang@cs.tu-berlin.de`

Sheet 10

Due: 7 July 2009

[Hints: Take the slide 12 in lecture 10 as basis for all three exercises.]

1. Consider the model $P(X,Y,Z) = P(X)\ P(Y|X)\ P(Z|X)$ with discrete random variables X,Y,Z. We are given a complete data set $\{(x_i, y_i, z_i)\}_{i=1}^N$. Prove that the maximum likelihood parameters (that maximize the *complete* data log-likelihood) of the model are

$$\pi_x \leftarrow \frac{c_x}{N}\ , \quad a_{yx} \leftarrow \frac{c_{yx}}{Nc_x}\ , \quad b_{zx} \leftarrow \frac{c_{zx}}{Nc_x} \tag{1}$$

where parameterization is $P(X=x) \equiv \pi_x$, $P(Y=y|X=x) \equiv a_{yx}$, $P(Z=z|X=x) \equiv b_{zx}$ and the counts are defined as

$$c_x = \sum_{i=1}^N [x_i = x] \tag{2}$$

$$c_{yx} = \sum_{i=1}^N [y_i = y][x_i = x] \tag{3}$$

$$c_{zx} = \sum_{i=1}^N [z_i = z][x_i = x] \tag{4}$$

[Hints: Take the function $L(\theta)$ in slide 12 as starting point. Take derivatives w.r.t. to parameters $\theta = (\boldsymbol{\pi}, \boldsymbol{a}, \boldsymbol{b})$, set them to zero with Lagrange multipliers ensuring the normalization of the parameter tables.]

2. Assume that X, Y, and Z are binary variables. Consider the same model, but X is unobserved. That is we have some data $\{(y_i, z_i)\}_{i=1}^N$. To simplify the notation, we define the counts

$$d_{yz} = \frac{1}{N}\ \sum_{i=1}^N [y_i = y][z_i = z]\ .$$

Derive the E-step and M-step of Expectation Maximization to learn the parameters of the model in terms of these counts. (In the M-step, parameter maximize the *expected* data log-likelihood.)

[Hints: The E-step is straight-forward. For the M-step, take the function $Q(\theta, \theta^{old})$ as starting point, express it in terms of the θ, θ^{old} and δ_*, take the derivative w.r.t. θ as above.]

3. Consider the same model. Analytically marginalize X to formulate the model $P(Y, Z; \theta)$ and thereby the *observed* data log-likelihood $\log P(Y, Z; \theta)$. Derive equations for the optimal parameters, that maximize this observed data log-likelihood.

[Hints: Take the function $\hat{L}(\theta)$ as starting point. Express it only in terms of parameters θ and relative counts d_*. Take derivatives as before.]