Machine Learning (lecture) Fall 2014: Exercise sheet 7

Return your results on paper in the Friday class on November 28. Please typeset and print or write legibly (we will downgrade for illegible handwriting).

You may team up in groups of two.

This sheet is rather lightweight. The problems essentially are practising exercises to get familiar with a clean probability notation. I invite you to invest the saved time in the optional "EM-digits" project which is announced separately, and which is awarded with course bonus points.

Problem 1 (25 points) The statement and derivation of the EM principle in the lecture notes (Section 10, downloadable as standalone pdf at the course homepage) was formulated in a version based on continuous-valued random variables, using pdf's. Your task: reformulate the essential parts of the EM principle for the case of discrete-valued random variables (which take values in finite observation spaces E). Concretely, give discrete-probability versions of equations 10.28 – 10.30, 10.34, 10.36. Assume that the hidden variable X take values in a finite space $H = \{h_1, ..., h_K\}$ and the observable variables Y in a space $A = \{a_1, ..., a_L\}$. Note that X and A usually are cross products of many random variables, and consequentially the spaces H and A are product spaces of very large (but finite) cardinality. The training data D is a single point in A! Be detailed when you write probability terms, always using the full-length notation $P(X = h, ...)$.

Problem 2 (25 points) Prove equation (10.19) in the lecture notes chapter 10 (referring to the version posted on the course homepage).

Problem 3 (50 points) The lecture notes in Section 10.7 give a procedure to compute $LL_{true} = \log P(D \mid \theta_{true})$, described after equation (10.45). Give a proof that this procedure actually does what it is supposed to do, namely, return LL_{true}.