STATISTICAL PRINCIPALS AND COMPUTATIONAL METHODS. 4^{th} SEMINAR – STATISTICAL LEARNING IN MRF-S

Exercise 1. Consider the Potts model

$$p(y) = \frac{1}{Z(\alpha)} \exp\left[-\alpha \sum_{ij\in\mathscr{E}} \delta(y_i \neq y_j)\right]$$

The only unknown parameter of the model is the Potts strength α. It should be learned according to the Maximum Likelihood principle from a training set of labelings $L = (y^1, y^2 \dots y^l)$.

a) Derive the gradient of the log-likelihood with respect to α.

b) Generalize the derivation for arbitrary *homogeneous* models, i.e. those where the pairwise functions ψ_{ij} are the same for all edges:

$$p(y) = \frac{1}{Z(\psi)} \exp\left[\sum_{ij\in\mathscr{E}} \psi(y_i, y_j)\right]$$

Exercise 2. Consider the probability distribution for pairs (x,y) that is decomposed into the prior and the conditional probability distribution, i.e.

$$p(x,y;\theta_1,\theta_2) = p(y;\theta_1)\cdot p(x|y;\theta_2)$$

Thereby each part depends on its own parameter θ_1 and θ_2 respectively.

a) Show that in the supervised case the parts (the corresponding parameters) can be learned independently.

b) Let the prior part be an MRF for labelings y (for example like the one from the previous exercise) and the conditional p.d. be conditionally independent, i.e.

$$p(x,y;\theta_1,\theta_2) = \frac{1}{Z(\theta_1)} \exp\big[-E(y;\theta_1)\big]\cdot \prod_i p(x_i|y_i;\theta_2)$$

Furthermore, assume that we are interested in the learning of the conditional part only, i.e. θ_1 is known. To be concrete let us consider a "non-parametric" probability distribution $p(x_i|y_i)$, i.e. it is given by a table, that assigns a value for each pair (x_i, y_i). For instance, if x_i represent pixel gray-values, the table is composed of K histograms of gray-values – one per label.

Show that in the supervised case the conditional probability distribution $p(x_i|y_i)$ can be learned exactly (globally optimal).