Final practice problems
ECE 271A
Department of Computer and Electrical Engineering
University of California, San Diego

Nuno Vasconcelos Fall 2008

1. Least squares with missing data Consider the least squares problem where we have two random variables Z and $\mathbf{X}$, such that

$$z = f(\mathbf{x}, \theta) + \epsilon, \tag{1}$$

where f is a polynomial with parameter vector θ

$$f(\mathbf{x}, \theta) = \sum_{k=0}^{K} \theta_k x^k \tag{2}$$

and ϵ a Gaussian random variable of zero mean and variance σ^2. As before, our goal is to obtain the ML estimate of the function, given an iid sample $\mathcal{D}$. However, due to a data gathering problem, some of the z_i values are missing. For simplicity, assume that $\mathcal{D} = \{(x_1, z_1), \dots, (x_m, z_m), x_{m+1}, \dots x_n\}$, i.e. the first m observations are complete but only the values of x_i are available for the remaining $n - m$ cases.

a) Write down the function to be minimized when the goal is to compute the ML estimate of the parameter vector θ with respect to $\mathcal{D}$ using standard ML procedures. What is the role of the x_i for which z_i is missing? Does this make sense?

b) Derive the EM equations for the ML estimation of the parameter vector θ.

c) Show that EM converges to the value of θ that maxmizes the cost function of **a**).

d) Can you think of a scenario in which it makes sense to use EM for this problem, instead of the direct maximization of **a**)?

2. ML estimation of multivariate t distribution A p dimensional random variable $\mathbf{W}$ has a multivariate t distribution with parameters μ, $\mathbf{\Sigma}$, and ν if, given the weight u,

$$P_{\mathbf{W}|U}(\mathbf{W}|u) = \mathcal{G}\left(\mathbf{W}, \mu, \frac{1}{u}\mathbf{\Sigma}\right)$$

where u has a Gamma distribution with parameters $(\frac{\nu}{2}, \frac{\nu}{2})$, i.e.

$$P_U(u; \nu) = \frac{\beta^{\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})} u^{\frac{\nu}{2}-1} e^{-\frac{\nu}{2}u}.$$

The weight u can be integrated to obtain the pdf

$$P_{\mathbf{W}}(\mathbf{w}; \mu, \mathbf{\Sigma}, \nu) = \frac{\Gamma\left(\frac{\nu+p}{2}\right) |\mathbf{\Sigma}|^{-1/2}}{(\pi\nu)^{p/2}\Gamma\left(\frac{\nu}{2}\right)\{1 + \frac{1}{\nu}(\mathbf{w}-\mu)^T\mathbf{\Sigma}^{-1}(\mathbf{w}-\mu)\}^{\frac{1}{2}(\nu+p)}}$$

but this looks nearly intractable, when the goal is to compute ML estimates. It is an example where, even though there is no missing data per se, the EM algorithm can still be very handy.

Given an iid sample $\mathcal{D} = \{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$, consider the problem of obtaining ML estimates for the parameters μ, and $\mathbf{\Sigma}$ (ν is assumed known). For this, consider the hidden variable U, a sequence of missing data $\mathcal{U} = \{u_1, \ldots, u_n\}$ drawn from it, and derive the steps of the EM algorithms that, considering ($\mathcal{D}$,$\mathcal{U}$) as complete data, leads to the set of parameters that maximize the likelihood of the incomplete data $\mathcal{D}$.