CS540 Spring 2010: homework 6

All references are to the 19feb10 version of my book. Also, please download pmtk 19feb10 version or newer.

1 Imputation using mixtures of Gaussians

The function `mixGaussImputationDemo` samples some data from a mixture of K Gaussians, "hides" some of the entries, fits a single Gaussian to the partially observed data using EM (sec 7.4), and then uses the estimated parameters to compute the posterior mean of the hidden variables (sec 7.3.1). This method only works well if $K = 1$ (see `gaussImputationDemo` for this special case), since modeling multi-modal data with 1 Gaussian does not work well. Your job is to improve this situation.

1. Derive the equations to compute $p(\mathbf{x}_h|\mathbf{x}_v, \boldsymbol{\theta})$ for a mixture model, where h are the hidden components and v are the visible components.
2. Implement your equations as a function called `mixGaussImpute`, with the same interface as `gaussImpute`. Apply your method to `mixGaussImputationDemo`, using the true (generating) parameters to do inference. Turn in new scatter plot and R^2 values.
3. Now derive the EM equations to fit a GMM (by MLE) when some components of the feature vector are missing, including the equations to compute the likelihood of a partially observed data vector.
4. Implement your equations as a function called `mixGaussMissingFitEm`, with a similar interface to `gaussFitMissingEm`. Also implement a function called `mixGaussLogprob`. You can use this to check that EM monotonically increases the observed data log-likelihood.
5. Apply your method to `mixGaussImputationDemo`, where you impute using the estimated parameters. Turn in your scatter plot and R^2 values.

2 EM for group lasso

Read 13.5.2–13.5.3, 13.6.4. Then modify `linregFitSparseEm` so it solves the group lasso problem. (You need to set the shape parameter for the Gamma prior according to the size of each group.) Apply your method to the data in `linregGroupLassoDemo`. Compare your results. (You should get the same answer; I don't know how the speed compares.)

3 Bayesian lasso

Read sec 13.5.2. Consider the following piece of the model

$$\tau_j^2 \sim \text{Exp}(\gamma^2/2) \quad (1)$$

$$w_j|\tau_j^2, \sigma^2 \sim \mathcal{N}(0, \sigma^2\tau_j^2) \quad (2)$$

Show that the posterior for τ_j^2 is an inverse Gaussian distribution

$$\frac{1}{\tau_j^2}|w_j, \sigma^2 \sim \text{InvGauss}(\frac{\gamma\sigma}{w_j}, \gamma^2) \tag{3}$$

4 LLA for SCAD

Read sec 13.6.6. Implement the LLA algorithm for optimizing linear regression with the SCAD regularizer. (See `scadPlot` for some handy functions.) Apply your method to the synthetic data used in `linregSparseEmSynthDemo` (be sure to use the same random number seed). Add your method to the list of existing methods and create new box-plots (as in fig 13.12) and plots of the weights. Turn in your code and figures.

5 Nearest shrunken centroids classifier

(Source: Ex 18.2 of [HTF09])

1. Show that the solution to 12.104 is given by 12.105–12.109, where s_j is the pooled standard deviation for feature j.
2. Implement the algorithm and apply it to the SRBCT dataset available from `http://www-stat.stanford.edu/~tibs/ElemStatLearn/data.html`. Try to reproduce figs 12.9–12.10 of my book. (See [HTF09, sec 18.2] for details. Recall that this book is available online for free at `http://www-stat.stanford.edu/~tibs/ElemStatLearn/download.html`.) Turn in your code and figures.

References

[HTF09] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer, 2009. 2nd edition.