Semidefinite Programming	**Lecture 27**
OR 6327 Spring 2012	**May 1, 2012**
Scribe: Daniel Fleischman	

1 The Sparse Covariance Selection Problem

Suppose we have n random variables and we want to detect their dependence on each other as well as estimating mean and covariance parameters.

Note that we may (wrongly) detect a dependence. Suppose variables X_1 and X_2 are (conditionally) independent of each other, but X_1 and X_3 are dependent, X_3 and X_4 are dependent and X_4 and X_2 are dependent. (Henceforth we'll use lower-case letters for random variables to avoid confusion with matrices.)

We will assume that the variables are multivariate normal parametrized by mean μ and covariance matrix Σ (Σ is pd). Then the density is $f(x) = c_1 \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$ for some normalizing constant c_1.

Let $y = (x_i; x_j)$ consist of two of the variables and let $K = \{1, 2, \ldots, n\} \setminus \{i, j\}$. Then the density of y given $x_K = \bar{x}_K$ is:

$$g(y) = c_2 \exp\left\{-\frac{1}{2}(y-\nu)^T\Sigma^{-1}_{\{i,j\},\{i,j\}}(y-\nu)\right\}$$

for some ν and c_2. Hence, x_i and x_j are independent given $x_K = \bar{x}_K$ if and only if $(\Sigma^{-1})_{ij} = 0$.

We sample N times to get $x_1, \ldots, x_N \in \mathbb{R}^n$ and compute the sample mean $\bar{x} = (x_1 + \ldots + x_N)/N$ and the sample variance $\bar{\Sigma} = ((x_1-\bar{x})(x_1-\bar{x})^T + \ldots + (x_N-\bar{x})(x_N-\bar{x})^T)/N$, and use maximum likelihood. The likelihood function given the data is:

$$h(\mu, \Sigma) := c_3 \det(\Sigma)^{-N/2} \prod_{j=1}^{N} \exp\left\{-\frac{1}{2}(x_j-\mu)^T\Sigma^{-1}(x_j-\mu)\right\}$$

for yet another normalizing constant c_3. For any Σ pd this is maximized over μ by $\mu = \bar{x}$. So we want to maximize:

$$\bar{h}(\Sigma) := h(\bar{x}, \Sigma) = c_3 \det(\Sigma)^{-N/2} \exp\left\{-\frac{1}{2}\Sigma^{-1} \bullet \bar{\Sigma}\right\} = \exp\left\{c_4 - \frac{N}{2}\text{lndet } \Sigma - \frac{N}{2}\Sigma^{-1} \bullet \bar{\Sigma}\right\}.$$

Since the exponential function is monotone, we want to maximize its argument. This is not concave in Σ. So we change variables. Let $X = \Sigma^{-1}$. Then we want to minimize $-\text{lndet } X + \bar{\Sigma} \bullet X$, which is convex.

Parametrizing our optimization by the inverse covariance has another advantage, since this is the matrix we expect (or hope) to be sparse. We will penalize the presence of nonzeros in X by adding an "l_1" term: $|||X|||_1 := ||\text{vec}(X)||_1$. So, at the end, we want to solve:

$$\min_{X \succ 0} \quad \underbrace{-\text{lndet } X + \bar{\Sigma} \bullet X}_{\text{smooth}} \quad + \quad \underbrace{\rho|||X|||_1.}_{\text{nice nonsmooth convex}}$$

We will use variable splitting and the alternating direction augmented Lagrangian method.

2 The Method

Reformulate as:

$$\begin{aligned} \min_{X \succ 0} \quad & -\text{lndet } X + \bar{\Sigma} \bullet X + \rho|||W|||_1 \\ \text{s.t.}: \quad & X - W = 0, \end{aligned}$$

and consider the augmented Lagrangian:

$$L(X, W, Y) := -\text{lndet } X + \bar{\Sigma} \bullet X + \rho|||W|||_1 - Y \bullet (X - W) + \frac{\beta}{2}||X - W||_F^2$$

for $\beta > 0$ fixed.

The idea is to start with $X_0 = W_0 = I$ and $Y_0 = 0$ and in each step do $(X_{k+1}, W_{k+1}) = \arg\min_{X,W} L(X, W, Y_k)$, and $Y_{k+1} := Y_k - \beta(X_{k+1} - W_{k+1})$.

But it is more efficient to minimize separately over X and W, to get:

- Start with $X_0 = W_0 = I$ and $Y_0 = 0$.
- In iteration k do:
  - $X_{k+1} = \arg\min_X L(X, W_k, Y_k)$
  - $W_{k+1} = \arg\min_W L(X_{k+1}, W, Y_k)$
  - $Y_{k+1} = Y_k - \beta(X_{k+1} - W_{k+1})$.

The two optimization steps are given by the following two propositions. First note that:

$$\begin{aligned} L(X, W, Y) &= -\text{lndet } X - (Y - \bar{\Sigma}) \bullet X + \frac{\beta}{2}||X - W||_F^2 + g(W, Y) \\ &= -\text{lndet } X + \frac{\beta}{2}||X - (W + \frac{1}{\beta}(Y - \bar{\Sigma}))||_F^2 + g'(W, Y) \end{aligned}$$

where $g(W, Y)$ and $g'(W, Y)$ are functions independent of X.

Proposition 1 *Let $B := W_k + \frac{1}{\beta}(Y_k - \bar{\Sigma})$ have eigenvalue decomposition $Q\Lambda Q^T$. Then the minimizer of $-\text{lndet } X + \frac{\beta}{2}||X - B||_F^2$ is $X = QMQ^T$ where $M = Diag(\mu)$ and $\mu_j = \frac{1}{2}\left(\lambda_j + \sqrt{\lambda_j^2 + 4\beta^{-1}}\right)$ for all j.*

Proof: The objective function is convex and smooth, with derivative $-X^{-1} + \beta(X - B)$, so it is minimized when $\beta X - X^{-1} = \beta B$.

Let $M = Q^T X Q$; then by premultiplying by Q^T and postmultiplying by Q we get $\beta M - M^{-1} = \beta\Lambda$. This is solved by a diagonal matrix $M = \text{Diag}\,(\mu)$ with $\beta\mu_j - \mu_j^{-1} = \beta\lambda_j$, for all j, with roots as given above. $\square$

Now, for the minimization problem over W, note that:

$$\begin{aligned} L(X, W, Y) &= -Y \bullet (X - W) + \rho|||W|||_1 + \frac{\beta}{2}||X - W||_F^2 + q(X, Y) \\ &= \rho|||W|||_1 + \frac{\beta}{2}||W - (X - \frac{1}{\beta}Y)||_F^2 + q'(X, Y) \end{aligned}$$

for $q(X, Y)$, $q'(X, Y)$ independent of W.

So we have:

Proposition 2 *Let $C = \beta X_{k+1} - Y_k$. Then the minimizer of $\rho|||W|||_1 + \frac{\beta}{2}||W - \frac{1}{\beta}C||_F^2$ is $W = \frac{1}{\beta}(C - P_{B_\infty^\rho}(C))$, where $B_\infty^\rho = \{Z \in \mathbb{M}^n : |z_{ij}| \leq \rho\, \forall i, j\}$ and $P_{B_\infty^\rho}$ is the projection on that set.*

Proof: Note that $C - P_{B_\infty^\rho}(C) = D$, where d_{ij} is 0 if $-\rho \leq c_{ij} \leq \rho$, is $c_{ij} + \rho$ if $c_{ij} < -\rho$ and $c_{ij} - \rho$ if $c_{ij} > \rho$.

The objective function separates completely over each entry of W and C, so we want:

$$\min_{w_{ij}} \rho|w_{ij}| + \frac{\beta}{2}(w_{ij} - \frac{1}{\beta}c_{ij})^2.$$

This is convex, and its subdifferential is:

$$\partial f(w_{ij}) = \begin{cases} \{\rho + (\beta w_{ij} - c_{ij})\} & \text{if } w_{ij} > 0, \\ \{-\rho + (\beta w_{ij} - c_{ij})\} & \text{if } w_{ij} < 0, \\ [-\rho, \rho] + \{\beta w_{ij} - c_{ij}\} & \text{if } w_{ij} = 0. \end{cases}$$

In all cases we see that W as in the statement of the proposition has entry w_{ij} with a subgradient of 0. $\square$