E9 203 Compressed Sensing and Sparse SP

29 Mar. 2021.

Last time: SBL; will continue today!

$$y = Ax + w, \quad x \in \mathbb{R}^N, \; y \in \mathbb{R}^m.$$

$$w \sim \mathcal{N}(0, \sigma^2 I_m) \;;\; x \sim \mathcal{N}(0, \Gamma)$$

$$\Rightarrow y = \mathcal{N}(0, \Sigma_y) \;;\; \Sigma_y = \sigma^2 I_m + A\Gamma A^T.$$

Choose Γ to max. $\log p(y; \Gamma, \sigma^2)$

$$\Rightarrow \min_{\Gamma} L(\Gamma) \triangleq \underline{-\log p(y;\Gamma)} = \log\det\Sigma_y + y^T\Sigma_y^{-1}y.$$

$$-L(\Gamma) \underset{\uparrow}{\geq} \mathbb{E}_{x|y;\Gamma^{(k)},\sigma^2}\left\{\log \frac{p(y,x;\Gamma,\sigma^2)}{p(x|y;\Gamma^{(k)},\sigma^2)}\right\}$$

Jensen's ineq.

$$\arg\max_{\Gamma} \mathbb{E}_{x|y;\Gamma^{(k)},\sigma^2}\{\log\det\Gamma - x^T\boxed{\Gamma^{-1}}x\}$$

$\max_{\Gamma}$ of the RHS led to:

$$\max_{\Gamma} -\left\{\sum_{i=1}^N \underline{\log\gamma_i} + \underline{\gamma_i^{-1}}\underbrace{\mathbb{E}_{x|y;\Gamma^{(k)},\sigma^2}\{x_i^2\}}_{\mu_i^2 + \Sigma_{w,ii}}\right\}$$

where $\mu = \sigma^{-2}\Sigma_w A^T y$; $\Sigma_w = (\sigma^{-2}A^TA + \Gamma^{(k)^{-1}})^{-1}$.

$$\Rightarrow \boxed{\gamma_i^{opt} = \mu_i^2 + \Sigma_{w,ii}.}$$

Thus, recipe for $\boxed{\max_{\Gamma} \log p(y;\Gamma,\sigma^2)}$:

Init: $\Gamma^{(0)} = I$; $k = 0$.

Repeat:

1. $q(x) = p(x|y;\Gamma^{(k)},\sigma^2)$
   $= \mathcal{N}(\mu, \Sigma_w)$
   $\left[\begin{array}{l}\mu = \sigma^{-2}\Sigma_w A^T y \\ \Sigma_w = \underbrace{(\sigma^{-2}A^TA + \Gamma^{(k)^{-1}})^{-1}}_{N\times N}.\end{array}\right.$ — E-STEP.

2. Optimize:
   $\gamma_i^{(k+1)} = \mu_i^2 + \Sigma_{w,ii}$, $i = 1, 2, \ldots, N.$ — M-STEP.

Until $L(\Gamma) = \log\det\Sigma_y + y^T\Sigma_y^{-1}y$, $\Sigma_y = (\sigma^2 I + A\Gamma A^T)$ converges

Output μ.

Remark 1: $\Sigma_w = (\sigma^{-2}A^TA + \Gamma^{(k)^{-1}})^{-1}$: inversion of an $N\times N$ matrix.

Matrix inversion lemma:

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}$$

We get

$$\boxed{\Sigma_w = \Gamma - \Gamma A^T(\underbrace{\sigma^2 I + A\Gamma A^T}_{m\times m})^{-1}A\Gamma}.$$

Remark 2: $\sigma^2 \to 0$

$$\begin{aligned}\mu &= \sigma^{-2}\Sigma_w A^T y \\ &= \sigma^{-2}(\sigma^{-2}A^TA + \Gamma^{-1})^{-1}A^Ty \\ &= (A^TA + \sigma^2\Gamma^{-1})^{-1}A^Ty \\ &= \Gamma^{\frac{1}{2}}(\Gamma^{\frac{1}{2}}A^T A\Gamma^{\frac{1}{2}} + \sigma^2 I)^{-1}\Gamma^{\frac{1}{2}}A^Ty\end{aligned}$$

Limit relation (see Wikipedia: pseudoinverse)

$$B^{\dagger} = \lim_{\delta\to 0}\underbrace{(B^TB + \delta I)^{-1}B^T} = \lim_{\delta\to 0}\underbrace{B^T(BB^T + \delta I)^{-1}}$$

$$\mu \underset{\sigma^2\to 0}{=} \Gamma^{\frac{1}{2}}(A\Gamma^{\frac{1}{2}})^{\dagger}y.$$

$$\begin{aligned}\Sigma_w &= \Gamma - \Gamma A^T(\sigma^2 I + A\Gamma A^T)^{-1}A\Gamma \\ &= \Gamma - \Gamma^{\frac{1}{2}}\,\Gamma^{\frac{1}{2}}A^T(\sigma^2 I + A\Gamma^{\frac{1}{2}}\Gamma^{\frac{1}{2}}A^T)^{-1}A\Gamma \\ &= \Gamma - \Gamma^{\frac{1}{2}}(A\Gamma^{\frac{1}{2}})^{\dagger}A\Gamma.\end{aligned}$$

Thm. Let $\sigma^2 = 0$, and let x_0 be the maximally sparse soln. to $y = Ax$. Assume $|\mathrm{supp}(x_0)| < m$, and $\|x_0\|_2 < \infty$. Let γ_0 (or Γ_0) be a vec. of prior variances s.t. $\boxed{x_0 = \Gamma_0^{\frac{1}{2}}(A\Gamma_0^{\frac{1}{2}})^{\dagger}y}$ where the min nonzero entry of Γ_0 is $> \delta > 0$, and $|\mathrm{supp}\,\gamma_0| = |\mathrm{supp}(x_0)|$. Then, the global min. of

$$\mathcal{L} = \log\det(\underbrace{\sigma^2 I + A\Gamma A^T}_{\Sigma_y}) + y^T(\sigma^2 I + A\Gamma A^T)^{-1}y$$

occurs at $\Gamma = \Gamma_0$ and $\sigma^2 = 0$.

Proof: We will s.t. $L \to -\infty$ at $\sigma^2 = 0$, $\Gamma = \Gamma_0$.

$$\log\det\Sigma_y \to -\infty, \quad y^T\Sigma_y^{-1}y \to \text{bounded}.$$

Now, when $\sigma^2 = 0$, $\Gamma = \Gamma_0$,

$$\mathrm{rank}(\Sigma_y) = \mathrm{rank}(A\Gamma_0A^T) \leq \mathrm{rank}(\Gamma_0) < m.$$

$$\Rightarrow \det\Sigma_y = 0, \quad \log\det\Sigma_y \to -\infty.$$

With $\sigma^2 = 0$, $\Gamma = \Gamma_0$, $y^T(A\Gamma_0A^T)^{-1}y \underset{(HW)}{=} x_0^T\Gamma_0^{\dagger}x_0 \leq \frac{1}{\delta}\|x_0\|_2^2 < \infty.$

$$\Rightarrow L \to -\infty \text{ at } \sigma^2 = 0, \; \Gamma = \Gamma_0. \qquad \square$$

Facts: (See D. Wipf & B. Rao's papers)

1. If A has the unique representation property (any subset of m cols of A are LI), then, the global min. of L can only occur at a degenerate γ, that produces a degenerate sparse soln, i.e., $|\mathrm{supp}(x)| < m$. (x is obtained from the posterior mean computed using the γ that attains the global min. of L)
2. Under mild assumps on A, there will exist no other degenerate sparse solns.
3. Can s.t. every local min. of L occurs at a sparse soln, i.e., $\|x\|_0 < m$, even in the noisy case.

Updating σ^2

In Step 2, must max. wrt σ^2

$$\mathbb{E}_{x|y;\Gamma^{(k)},\sigma^{(k)^2}}\left\{\log\frac{p(y,x;\Gamma^{(k)},\sigma^2)}{p(x|y;\Gamma^{(k)},\sigma^{(k)^2})}\right\}$$

Isolating the terms that dep. only on σ^2,

$$\mathbb{E}_{\underset{\circledast}{\boxed{x|y,\Gamma^{(k)},\sigma^{(k)^2}}}}\{\log p(y|x;\sigma^2)\}$$

$$= \mathbb{E}_{\circledast}\left\{-\frac{\|y - Ax\|_2^2}{2\sigma^2} - \frac{m}{2}\log\sigma^2\right\}$$

Diff. w.r.t. σ^2, set $= 0$:

$$\sigma^2 = \frac{1}{m}\mathbb{E}_{\circledast}\{\|\underbrace{y - Ax}_{y - A\mu + A(\mu - x)}\|_2^2\}$$

$$(\mu - x)^TA^TA(\mu - x) = \mathrm{tr}(A^TA(\mu - x)(\mu - x)^T)$$

$$= \frac{1}{m}\mathbb{E}_{\circledast}\{\|y - A\mu\|_2^2 + \mathrm{tr}(A^TA(\mu - x)(\mu - x)^T)\}$$

$$= \frac{1}{m}\{\|y - A\mu\|_2^2 + \mathrm{tr}(A^TA\Sigma_w)\}$$

$$\Sigma_\omega = \left(\sigma^{(k)^{-2}} A^T A + \Gamma^{(k)^{-1}} \right)^{-1}$$

$$\Rightarrow \left(\Sigma_\omega^{-1} - \Gamma^{(k)^{-1}} \right) = \sigma^{(k)^{-2}} A^T A$$

$$\Rightarrow A^T A = \underline{\underline{\sigma^{(k)^{2}}}} \underline{\left(\Sigma_\omega^{-1} - \Gamma^{(k)^{-1}} \right)}$$

$$= \frac{1}{m} \left\{ \| y - A\mu \|_2^2 + \sigma^{(k)^2} \operatorname{tr} \left(\left(\Sigma_\omega^{-1} - \Gamma^{(k)^{-1}} \right) \Sigma_\omega \right) \right\}$$

$$= \frac{1}{m} \left\{ \| y - A\mu \|_2^2 + \sigma^{(k)^2} \operatorname{tr} \left(I - \underline{\underline{\Gamma^{(k)^{-1}}}} \Sigma_\omega \right) \right\}$$

$$\boxed{\sigma^2 = \frac{1}{m} \left\{ \| y - A\mu \|_2^2 + \sigma^{(k)^2} \sum_{i=1}^{N} \left(1 - \gamma_i^{(k)^{-1}} \Sigma_{\omega, ii} \right) \right\}}$$

Q. Why does SBL lead to sparse solns. ?

Suppose we place a "hyper prior" on γ:

Inverse gamma distbn :

$$p_\gamma(\gamma) = \frac{\beta^\alpha}{\Gamma(\alpha)} \gamma^{-\alpha-1} e^{-\beta/\gamma}, \quad \gamma > 0, \ \underline{\alpha, \beta > 0}.$$

$$p_x(x) = \prod_{i=1}^{N} \int \frac{1}{\sqrt{2\pi\gamma_i}} e^{-\frac{x_i^2}{\gamma_i}} \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{1}{\gamma_i^{\alpha+1}} e^{-\beta/\gamma_i} d\gamma_i$$

$z_i = 1/\gamma_i$ and simplify

$$= \prod_{i=1}^{N} \int \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{x_i^2}{2} + \beta\right) z_i} \cdot z_i^{\alpha - \frac{1}{2}} dz_i$$

$$\propto \prod_{i=1}^{N} \left(\frac{x_i^2}{2} + \beta \right)^{-\left(\alpha + \frac{1}{2}\right)}.$$

As, $\alpha, \beta \to 0$, this becomes

$$\propto \prod_{i=1}^{N} \frac{1}{\underline{\underline{|x_i|}}}.$$ Clearly encourages sparsity. $\underline{\underline{|x|^p,\ p < 1}}$