E9 203 Compressed Sensing & Sparse SP.

31 Mar. 2021.

Last time:

- SBL algo.
- Speed-up via matrix inversion lemma
- Analysis of SBL
- Updating σ^2 $\qquad y = Ax + w, \quad w \sim \mathcal{N}(0, \sigma^2 I)$
- Hyperprior on γ:

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

$$\Rightarrow \quad p_x(x) = \int p(x|\Gamma)\, p(\Gamma)\, d\Gamma$$
$$\propto \prod_{i=1}^{N} \left(b + \frac{x_i^2}{2}\right)^{-(a+\frac{1}{2})}$$

as $a, b \to 0$, this becomes

$$\underline{\propto \prod_{i=1}^{N} \frac{1}{|x_i|}} .$$

$$\Rightarrow \max_x p(x|y) = \max_x \frac{\overbrace{p(y|x)}^{\text{Gaussian}} \overbrace{p(x)}^{\propto \frac{1}{|x_i|}}}{p(y)}$$

$$\Rightarrow \min_x -\log p(y|x) - \log p(x)$$

$$\Rightarrow \min_x \underline{\underline{\frac{\|y - Ax\|_2^2}{2\sigma^2} + \sum_{i=1}^{N} \log |x_i|}} \quad : \text{promotes sparsity}$$

Today:

- Further discⁿ on the inverse gamma hyperprior
- Reweighted ℓ_1 and ℓ_2 algos for SBL.

Idea: Instead of working with $\underline{\underline{p_x(x) \propto \prod_{i=1}^{N} \left(b + \frac{x_i^2}{2}\right)^{-(a+\frac{1}{2})}}}$,

we define an approx. to $p_x(x)$.

In fact: we consider a family of approx., and pick the member of the family that max. Bayesian evidence

$\Rightarrow$ Link betⁿ $p_x(x)$ and $p_x(x; \Gamma_{ML})$.

Let $\underline{z_i = x_i^2}$, and define

$$\underline{f(z_i) = \log p_x(x_i)} = \underbrace{-\left(a + \tfrac{1}{2}\right) \log\left(b + \tfrac{z_i}{2}\right)}_{\text{Convex fn. of } z_i} + \text{const.}$$

Any convex fn. $f(z)$ can be written in its dual form:

$$\boxed{\begin{aligned} f(z) &= \sup_\lambda \ \lambda z - f^*(\lambda) \\ f^*(\lambda) &\triangleq \sup_z \ \lambda z - f(z) \end{aligned}} \quad \text{convex conjugate fn.}$$

$$\underline{f^*(\lambda)} = \sup_z \ \lambda z + \left(a + \tfrac{1}{2}\right) \log\left(b + \tfrac{z}{2}\right)$$

$$\Rightarrow \lambda + \frac{(a + \frac{1}{2})}{(b + \frac{z}{2})} \cdot \frac{1}{2} = 0 \Rightarrow \frac{a + \frac{1}{2}}{b + \frac{z}{2}} = -2\lambda$$

$$\Rightarrow \underline{z} = -2b - \frac{(a + \frac{1}{2})}{\lambda} .$$

$$\underline{\underline{f^*(\lambda)}} = -2b\lambda - \left(a + \tfrac{1}{2}\right) + \left(a + \tfrac{1}{2}\right) \log\left(\frac{-(a + \frac{1}{2})}{2\lambda}\right)$$

$$f(z) = \max_\lambda \ \lambda z - f^*(\lambda)$$
$$= \max_\lambda \ \lambda z + 2b\lambda - \left(a + \tfrac{1}{2}\right) \log\left(-\frac{(a + \frac{1}{2})}{2\lambda}\right)$$

$\underline{\underline{\lambda \leq 0}}$.

Replace λ with $-\frac{1}{2\gamma}$,

$$\log p_x(x) \propto \max_{\gamma \geq 0} -\frac{z}{2\gamma} - \frac{b}{\gamma} - \left(a + \tfrac{1}{2}\right)\left(\underline{\log \gamma} + \underline{\underline{\log\left(a + \tfrac{1}{2}\right)}}\right)$$

$$p_x(x) \propto \max_{\gamma \geq 0} e^{-\frac{x^2}{2\gamma}} e^{-\frac{b}{\gamma}} \gamma^{-(a+\frac{1}{2})}$$

as $a, b \to 0$,

$$\underline{\underline{p_x(x)}} \propto \max_{\gamma \geq 0} \underline{\underline{\frac{1}{\sqrt{\gamma}} e^{-\frac{x^2}{2\gamma}}}}, \quad \text{Gaussian distbn!}$$

In SBL, we use a lower bound on $p_x(x)$ by letting γ be a "hyperparameter" and use type II ML to find a "tight" lower bound

The prior: $p_x(x) \propto \prod_{i=1}^{N} \left(b + \frac{x_i^2}{2}\right)^{-(a+\frac{1}{2})}$

encourages sparsity as $a, b \to 0$, but direct MAP estⁿ is hard

So, using the dual form representation, we get

$$p_x(x) \propto \prod_{i=1}^{N} \max_{\gamma_i} \frac{1}{\sqrt{2\pi\gamma_i}} \gamma_i^{-a} e^{-\frac{b}{\gamma_i}} e^{-\frac{x_i^2}{2\gamma_i}}$$

as $a, b \to 0$, this becomes

$$p_x(x) \propto \prod_{i=1}^{N} \max_{\gamma_i \geq 0} \underbrace{\frac{1}{\sqrt{2\pi\gamma_i}} e^{-\frac{x_i^2}{2\gamma_i}}}_{\substack{\text{Parametrized set of approx.} \\ \text{to } p_x(x). \quad \mathcal{N}(0, \Gamma).}}$$

Mckay update to speed up SBL:

$$L(\gamma) = -\log p(y; \gamma)$$
$$= \log \det\left(\sigma^2 I + A\Gamma A^T\right) + y^T \left(\sigma^2 I + A\Gamma A^T\right)^{-1} y$$

Differentiating w.r.t. γ_i and setting $= 0$, we arrive at the fixed pt. update:

$$\boxed{\gamma_i = \frac{\mu_i^2}{1 - \gamma_i^{-1} \Sigma_{ii}}} \qquad \Big| \ \text{HW}$$

where

$$\left[\begin{aligned} \mu &= \Gamma A^T \Sigma_y^{-1} y = \mathbb{E}\{x | y; \Gamma\} \\ \Sigma &= \text{Cov}\{x | y; \Gamma\} = \Gamma - \Gamma A^T \Sigma_y^{-1} A \Gamma \end{aligned}\right.$$

Considerably faster than EM. [Next time!]

Reweighted ℓ_1 min. approach for SBL cost fn. min.

$$\mathcal{L}(\gamma) = \underbrace{\log \det\left(\sigma^2 I + A\Gamma A^T\right)}_{\text{Concave in } \Gamma} + \underbrace{y^T \left(\sigma^2 I + A\Gamma A^T\right)^{-1} y}_{\text{Convex in } \Gamma.}$$

$$\underline{\log \det \Sigma_y} = \min_z \ z^T \gamma - g^*(z)$$

$$g^*(z) \triangleq \min_\gamma \ z^T \gamma - \log \det \Sigma_y$$

Thus, $\mathcal{L}(\gamma, z) = \underline{z^T \gamma - g^*(z) + y^T \left(\sigma^2 I + A\Gamma A^T\right)^{-1} y}$

$$\geq L(\gamma) .$$

Following the MM approach, we min the upper bd.

For a given z, we $\min_\gamma L(\gamma, z)$

For a given γ, we use the computation of $g^*(z)$ to update z: from the defn. of $g^*(z)$,

$$\underline{\underline{z_{opt}}} = \nabla_\gamma \log \det \Sigma_y .$$

Thus, we have the foll. procedure:

Init: z_i = something convenient, e.g. $z_i = 1$

Repeat:

$\gamma \leftarrow \underline{\arg\min_\gamma L_z(\gamma)} = \underline{z^T \gamma + y^T \Sigma_y^{-1} y}$ — ①

$z \leftarrow \nabla_\gamma \log \det \Sigma_y$ — ②

Until convergence of γ to γ_*

$\mathbb{E}\{x | y; \Gamma\} = \Gamma_* A^T \Sigma_{y*}^{-1} y$

Output $x_* = \dots$

How do we compute ①?

$\Rightarrow$ Weighted ℓ_1!

$$y^T \Sigma_y^{-1} y = \min_{x} \frac{1}{\sigma^2} \|y - Ax\|_2^2 + \sum_{i=1}^{N} \frac{x_i^2}{\gamma_i} \qquad [HW]$$

$$x_* = \arg\min_{x} \|y - Ax\|_2^2 + 2\sigma^2 \sum_{i=1}^{N} \sqrt{z_i}\, |x_i|$$

and then set

$$\gamma_i = (\sqrt{z_i})^{-1} |x_{*,i}|, \quad i = 1, 2, \dots, N.$$

An auxiliary upper bound

$$L_z(\gamma, x) = \sum_{i=1}^{N} z_i \gamma_i + \frac{x_i^2}{\gamma_i} + \frac{1}{\sigma^2} \|y - Ax\|_2^2 \geq L_z(\gamma)$$

Jointly convex in γ, x.

Given x, $\gamma_i = z_i^{-1/2} |x_i|$ is optimal.

Substituting,

$$L_z(\gamma_{opt}, x) = \sum_{i=1}^{N} \sqrt{z_i}\, |x_i| + \sqrt{z_i}\, |x_i| + \frac{1}{\sigma^2} \|y - Ax\|_2^2$$

$\Rightarrow$ ① can be solved by finding

$$x_* = \arg\min_{x} \|y - Ax\|_2^2 + 2\sigma^2 \sum_{i=1}^{N} \sqrt{z_i}\, |x_i|$$

And then setting

$$\gamma_i = z_i^{-\frac{1}{2}} |x_{*,i}|.$$

Global Convergence:

Thm. From any initialization $\Gamma^{(0)}$, the seq. of hyperparameters generated via ① and ② is guaranteed to converge monotonically to a local min or saddle pt. of the SBL cost fn. $L(\gamma)$.

Proof: See Wipf and Nagarajan, "A new view of automatic relevance determination (ARD)".