E9 203 Compressed Sensing & Sparse SP

26 Mar. 2021.

Last time:
- Analysis of reweighted algos
  - convergence.

Today:
- Sparse Bayesian learning

Sparse Bayesian learning:

Model: $y = Ax + w$; $y \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times N}$.

$w \sim \mathcal{N}(0, \sigma^2 I_m)$ → $\boxed{\sigma^2 \text{ known}}$ / σ^2 unknown.

$$p(y|x;\sigma^2) = \frac{1}{\sqrt{(2\pi\sigma^2)^m}} \exp\left(-\frac{\|y-Ax\|^2}{2\sigma^2}\right).$$

Need a prior on x. Choose:

$$\boxed{p(x;\gamma) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\gamma_i}} \exp\left(-\frac{x_i^2}{2\gamma_i}\right)} \leftarrow$$

$\gamma = [\gamma_1, \gamma_2, \ldots, \gamma_N]^T$ hyperparameters

$x \sim \mathcal{N}(0, \Gamma)$, $\Gamma = \mathrm{diag}([\gamma_1, \ldots, \gamma_N])$

$$\mathbb{E}\{yy^T\} = \mathbb{E}\{(Ax+w)(Ax+w)^T\} = \mathbb{E}\{Axx^TA^T + ww^T + \underbrace{Axw^T + wx^TA^T}_{0}\}$$

$$= \boxed{A\Gamma A^T + \sigma^2 I_m \triangleq \Sigma_y}$$

$y \sim \mathcal{N}(0, \Sigma_y)$

If Γ is known, can recover x via the conditional mean: $\mathbb{E}\{x|y\}$.

$$p(x|y;\Gamma,\sigma^2) = \frac{p(y|x;\Gamma,\sigma^2)\,p(x;\Gamma)}{p(y;\Gamma,\sigma^2)}$$

$$= \frac{\text{Gaussian} \times \text{Gaussian}}{\text{Gaussian}} = \text{Gaussian}$$

$$= \frac{1}{\sqrt{(2\pi)^N \det \Sigma_\omega}} \exp\left(-\frac{(x-\mu)^T \Sigma_\omega^{-1} (x-\mu)}{2}\right)$$

$\Sigma_\omega = (\sigma^{-2}A^TA + \Gamma^{-1})^{-1}$; $\boxed{\mu = \sigma^{-2}\Sigma_\omega A^T y}$

$\mathbb{E}\{x|y\}$; MAP/MMSE estimate of x.

$\boxed{\hat{x} = \mu}$

Property: $\mathrm{supp}(\hat{x}) = \mathrm{supp}(\gamma)$.

$$\boxed{p(x|y;\Gamma,\sigma^2) = \mathcal{N}(\mu, \Sigma_\omega)}$$

Σ_ω ← Not covariance of ω!

$$\boxed{\begin{aligned}\Sigma_\omega &= (\sigma^{-2}A^TA + \Gamma^{-1})^{-1} \\ \mu &= \sigma^{-2}\Sigma_\omega A^T y\end{aligned}}$$

$\mathbb{E}\{x_i^2\} = \mu_i^2 + \Sigma_{\omega,ii}$

But we don't know Γ: So, replace $p(x|y;\Gamma,\sigma^2)$ with $p(x|y,\hat{\Gamma},\sigma^2)$ where $\hat{\Gamma}$ is the ML est. of Γ from y. $\boxed{\text{Type II ML procedure}}$

$$\hat{\Gamma} = \arg\max_{\Gamma} p(y;\Gamma,\sigma^2) \longrightarrow \mathcal{N}(0, \underbrace{\Sigma_y}_{(A\Gamma A^T + \sigma^2 I_m)})$$

$$= \arg\max_{\Gamma} \log p(y;\Gamma,\sigma^2)$$

$$\log p(y;\Gamma,\sigma^2) = \log\left(\frac{1}{\sqrt{(2\pi)^m \det\Sigma_y}} \exp\left(-\frac{y^T\Sigma_y^{-1}y}{2}\right)\right)$$

$$= -\frac{y^T\Sigma_y^{-1}y}{2} - \frac{1}{2}\log\det\Sigma_y - \frac{m}{2}\log(2\pi)$$

Want

$$\arg\min_{\Gamma} L = \underbrace{\log\det\Sigma_y}_{\text{concave in }\Gamma} + \underbrace{y^T\Sigma_y^{-1}y}_{\text{convex in }\Gamma}$$

$$\Sigma_y = A\Gamma A^T + \sigma^2 I_m$$

We will use majorization maximization.

Step 1: Find $L(\Gamma|\Gamma^{(k)})$ s.t.

(1) $L(\Gamma) \le L(\Gamma|\Gamma^{(k)})$ $\forall \Gamma$

(2) $L(\Gamma^{(k)}) = L(\Gamma^{(k)}|\Gamma^{(k)})$

(3) $L(\Gamma|\Gamma^{(k)})$ is convenient for optimization.

Step 2: Use the iterative algo:

Init: $\Gamma^{(0)}$: something convenient

Repeat: $k = 0, 1, \ldots$
1. Form $L(\Gamma|\Gamma^{(k)})$
2. Solve $\Gamma^{(k+1)} = \arg\min_{\Gamma} L(\Gamma|\Gamma^{(k)})$
3. Set $k \leftarrow k+1$

Until convergence.

For step 1, note that

$$\log p(y;\Gamma,\sigma^2) = \log\int p(y,x;\Gamma,\sigma^2)\,dx \leftarrow$$

$-L(\Gamma)$, so we want a lower bound on the LHS.

$$= \log\int q_x(x) \boxed{\frac{p(y,x;\Gamma,\sigma^2)}{q_x(x)}}\,dx$$

$\boxed{q_x(x)}$ is any arbitrary distr. on x.

By Jensen's ineq.,

$$\log p(y;\Gamma,\sigma^2) \ge \int q_x(x) \log\left[\frac{p(y,x;\Gamma,\sigma^2)}{q_x(x)}\right] dx$$

(True for any $q_x(x)$ with the same support as $p(y,x;\Gamma,\sigma^2)$ — mild assump.)

Let's choose $q_x(x) = p(x|y;\Gamma^{(k)},\sigma^2)$

To complete Step 1, need t.s.t. the above ineq. becomes $=$ when $\Gamma = \Gamma^{(k)}$.

$$\text{RHS}: \int p(x|y;\Gamma^{(k)},\sigma^2) \log \underbrace{\frac{p(y,x;\Gamma^{(k)},\sigma^2)}{p(x|y;\Gamma^{(k)},\sigma^2)}}_{\left[\frac{p(y,x;\Gamma^{(k)},\sigma^2)}{p(y;\Gamma^{(k)},\sigma^2)}\right]} dx$$

$$\int p(x|y;\Gamma^{(k)},\sigma^2) \log p(y;\Gamma^{(k)},\sigma^2)\,dx$$

$$= \log p(y; \Gamma^{(k)}, \sigma^2) = \text{LHS}!$$

Thus, the choice $q_x(x) = p(x \mid y; \Gamma^{(k)}, \sigma^2)$ satisfies the cond$\underline{\text{ns}}$ of step 1.

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

$$\log \frac{p(y, x; \Gamma, \sigma^2)}{p(x \mid y; \Gamma^{(k)}, \sigma^2)} = \log \left(\frac{\cancel{p(y \mid x; \sigma^2)}\, p(x; \Gamma)}{\underbrace{\left[\frac{\cancel{p(y \mid x; \sigma^2)}\, p(x; \Gamma^{(k)})}{p(y; \Gamma^{(k)}, \sigma^2)} \right]}_{\text{Does not dep. on } \Gamma!}} \right)$$

$$= \log p(x; \Gamma) + \text{terms that dep. only on } \Gamma^{(k)}, \text{ not } \Gamma.$$

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

$$= \arg\max_{\Gamma} \mathbb{E}_{x \mid y; \Gamma^{(k)}, \sigma^2} \left\{ \underbrace{\log p(x; \Gamma)}_{-\log |\Gamma| - x^T \Gamma^{-1} x} \right\}$$

Since $\Gamma = \text{diag}([\gamma_1, \ldots \gamma_N]^T)$,

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

$$\log \gamma + \frac{1}{\gamma} c \rightarrow \frac{1}{\gamma} - \frac{1}{\gamma^2} c = 0 \text{ or } \boxed{\gamma = c}$$

$$\Rightarrow (\gamma_{opt})_i = \mu_i^2 + \Sigma_{\omega, ii}$$

Algo: Next time!