Machine learning under physical constraints
Unsupervised learning in DA

Sixin Zhang
(sixin.zhang@toulouse-inp.fr)

Outline

Unsupervised learning of Linear Dynamical systems

MLE approach and EM algorithm

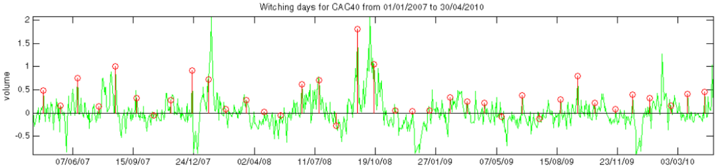
Outline

Unsupervised learning of Linear Dynamical systems

MLE approach and EM algorithm

A special ODS for financial time series

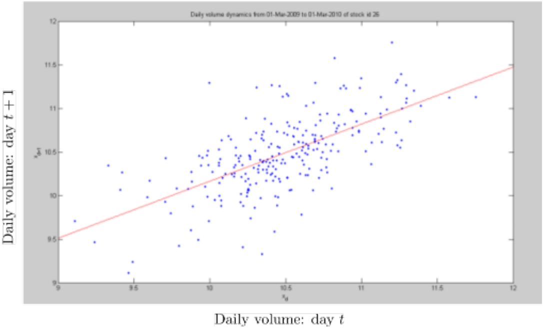
- Build a statistical model of financial time series to analyze stock volume changes under news impact



- Model dynamics using **Linear dynamical systems** (Kalman Filter).

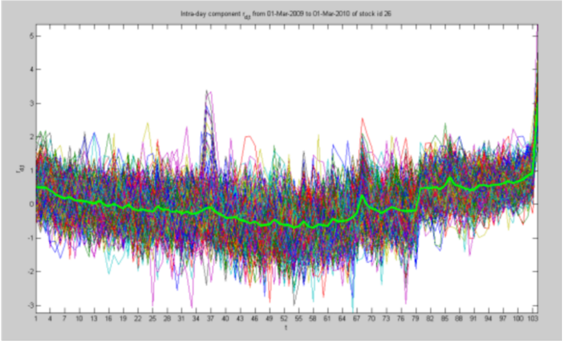
A special ODS for financial time series

- Model daily volume dynamics x_t over t.
- Assume $x_0 \sim \mathcal{N}(\mu_0, \Sigma_0), A \in \mathbb{R}, B \in \mathbb{R}, C > 0$
- Dynamics: $x_{t+1} | x_t \sim \mathcal{N}(Ax_t + B, C)$



A special ODS for financial time series

- Model within-day volume $y_t \in \mathbb{R}^d$ for day t,



- Assume $E \in \mathbb{R}^d$, F p.d matrix, $\mathbb{1}$ vector of ones

- Observation: $y_t | x_t \sim \mathcal{N}(\mathbb{1} x_t + E, F)$

Problem formulation

- Can we estimate the densities $p(x_t|x_{t-1})$ and $p(y_t|x_t)$ for $t = 1, \cdots, T$ from $y = (y_1, \cdots, y_T)$?
- Problem: estimate $\theta = (\mu_0, \Sigma_0, A, B, C, E, F)$ from y.
- Challenge: assume x_t is not observed (because certain y_t may be impacted by news, thus not observed)
- Idea: Maximum likelihood estimation with EM algorithm
- Reference: Pattern Recognition and Machine Learning by Christopher M. Bishop

Outline

Unsupervised learning of Linear Dynamical systems

MLE approach and EM algorithm

MLE approach and EM algorithm

- Identifiability condition (to restrict E):

$$E \perp \mathbb{1}$$

- Estimate θ from $y = (y_1, \cdots, y_T)$,

$$\max_{\theta} \log p(y|\theta)$$

- Estimating of the latent variables $x = (x_0, \cdots, x_T)$ makes the maximiation of θ easier,

$$\begin{aligned} \text{E step :} \quad & \max_{\phi} \int \log \frac{p(x|y,\theta)}{q(x|y,\phi)} q(x|y,\phi) dx \\ \text{M step :} \quad & \max_{\theta} \int \log p(x,y|\theta) q(x|y,\phi) dx \end{aligned}$$

MLE approach and EM algorithm

- KF smoother makes the E step possible: compute

$$p(x_t|y_1, \cdots, y_T, \theta)$$

$$p(x_t, x_{t+1}|y_1, \cdots, y_T, \theta)$$

- **Factorization** of $\log p$ makes the M step possible: in KF,

$$\log p(x, y|\theta) = \log p(x_0|\theta) + \sum_{t=0}^{T-1} \log p(x_{t+1}|x_t, \theta) + \sum_{t=1}^{T} \log p(y_t|x_t, \theta).$$

Road-map: 3 sub-problems in M step

- Decompose the M-step using factorization,

$$\theta^{k+1} \in \arg\max_{\theta} \int \log p(x, y|\theta) q(x|y, \phi^{(k)}) dx$$

- **Problem 1**: estimate (μ_0, Σ_0)

$$\max_{\theta} \int \log p(x_0|\theta) q(x_0|y, \phi^{(k)}) dx_0$$

  This requires to solve the **E step** for $q(x_0|y, \phi^{(k)})$ at θ^k.
- Problem 2: estimate (A, B, C)
- Problem 3: estimate (E, F)

Problem 1: analytical formula of M step

$$\ell_1 = \int \log p(x_0|\theta) q(x_0|y, \phi^{(k)}) dx_0$$

- Under $q(x_0|y, \phi^{(k)})$, x_0 follows $\mathcal{N}(\mu_{0|T}, \Sigma_{0|T})$
- Develop the log-likelihood,

$$-\ell_1 \propto \mathbb{E}_{x_0}((x_0 - \mu_0)^\intercal \Sigma_0^{-1} (x_0 - \mu_0)) + \log |\Sigma_0|$$

- To maximize ℓ_1, compute the critical points

$$\frac{\partial \ell_1}{\partial \mu_0} = 0, \frac{\partial \ell_1}{\partial \Sigma_0} = 0 \quad \Rightarrow \quad \mu_0^{(k)} = \mu_{0|T}, \Sigma_0^{(k)} = \Sigma_{0|T}$$

E step: How to compute $\mathcal{N}(\mu_{0|T}, \Sigma_{0|T})$?

- KF propogation and analysis: compute $p(x_T|y)$ recursively.
- KF smoother: a key to compute

$$x_{t+1}|y_1, \cdots, y_T \Rightarrow x_t|y_1, \cdots, y_T$$

- Analysis step: $x_t|y_1, \cdots, y_t \sim \mathcal{N}(\mu_{t|t}, \Sigma_{t|t})$
- Propagation step:

$$x_{t+1}|y_1, \cdots, y_t \sim \mathcal{N}(\mu_{t+1|t}, \Sigma_{t+1|t})$$

E step: KF smoother

- Smoother: given

$$x_{t+1}|y_1, \cdots, y_T \sim \mathcal{N}(\mu_{t+1|T}, \Sigma_{t+1|T})$$

compute

$$x_t|y_1, \cdots, y_T \sim \mathcal{N}(\mu_{t|T}, \Sigma_{t|T})$$

- We verify that

$$\mu_{t|T} = \mu_{t|t} + J_t(\mu_{t+1|T} - \mu_{t+1|t})$$

$$\Sigma_{t|T} = \Sigma_{t|t} + J_t(\Sigma_{t+1|T} - \Sigma_{t+1|t})J_t^\intercal$$

where $J_t = \Sigma_{t|t}^\intercal A^\intercal (C + A\Sigma_{t|t}A^\intercal)^{-1}$

E step: KF smoother (technical detail)

- Verify that $p(x_t|x_{t+1}, y_1, \cdots, y_T, \theta)$ follows

$$\mathcal{N}(\mu_{t|t+1|T}, \Sigma_{t|t+1|T})$$

where

$$\mu_{t|t+1|T} = \mu_{t|t} + J_t[x_{t+1} - (A\mu_{t|t} + B)]$$
$$\Sigma_{t|t+1|T} = \Sigma_{t|t} - J_t A \Sigma_{t|t}$$

- This implies the joint distribution (x_{t+1}, x_t) given y is

$$\mathcal{N}\left(\begin{pmatrix} \mu_{t+1|T} \\ \mu_{t|T} \end{pmatrix}, \begin{pmatrix} \Sigma_{t+1|T}, & (J_t \Sigma_{t+1|T})^\intercal \\ J_t \Sigma_{t+1|T}, & \Sigma_{t|T} \end{pmatrix}\right)$$

M step: analytical formula of Problem 2

- Consider $\ell_2 = \sum_t \ell_{2,t}$, where

$$\begin{aligned}\ell_{2,t} &= \int \log p(x_{t+1}|x_t, \theta) q(x_{t+1}, x_t | y, \phi^{(k)}) dx_{t+1} dx_t \\ &\propto -\mathbb{E}_{(x_{t+1}, x_t)}[(x_{t+1} - Ax_t - B)^\intercal C^{-1}(x_{t+1} - Ax_t - B)] - \log |C|\end{aligned}$$

- Compute critical points of ℓ_2 using matrix calculus,

$$\frac{\partial \ell_2}{\partial A} = 0, \quad \frac{\partial \ell_2}{\partial B} = 0, \quad \frac{\partial \ell_2}{\partial C} = 0$$

- Exercise: find out the optimal solution.

M step: analytical formula of Problem 3

- Consider $\ell_3 = \sum_t \ell_{3,t}$, where

$$\ell_{3,t} = \int \log p(y_t|x_t, \theta) q(x_t|y, \phi^{(k)}) dx_t$$
$$\propto -\mathbb{E}_{x_t}[(y_t - \mathbb{1}x_t - E)^\intercal F^{-1}(y_t - \mathbb{1}x_t - E)] - \log |F|$$

- Due to the constraint on E, we compute critical points of ℓ_3 using the Lagrangian method.
- Exercise: find out the optimal solution (check `https://cs.nyu.edu/~zsx/rapport_mva_sixin.pdf`).