Probabilistic Time Series Analysis: Lab 4

Tim Kunisky

September 26, 2018

Feedback on Lab 2 Assignment

Feedback on Lab 2 Assignment

- You don't need to answer the questions in the text (unless you want to), the only part I plan to grade is what you fill in for the code blocks.
- Several people made mistakes generating $y_t \approx x_{t-5}$, instead generating $y_t \approx x_{t+5}$ or $y_t \approx x_t$—look at your plots!

Feedback on Lab 2 Assignment

- You don't need to answer the questions in the text (unless you want to), the only part I plan to grade is what you fill in for the code blocks.
- Several people made mistakes generating $y_t \approx x_{t-5}$, instead generating $y_t \approx x_{t+5}$ or $y_t \approx x_t$—look at your plots!

Notes on coding:

- If possible, avoid building or manipulating NumPy arrays with loops.
- In serious numerical work, when solving $\boldsymbol{Ax} = \boldsymbol{b}$,

  NEVER use `np.dot(np.linalg.inv(A), b)`.

  ALWAYS use `np.linalg.solve(A, b)`, or `scipy.linalg.lu_factor` and `scipy.linalg.lu_solve`.

Today's Topic

Review of the EM algorithm, and correcting some vague comments and mistakes from last session.

LDS Model and Learning Task

$$\begin{array}{ccccccccc}
\cdots & \to & z_{t-2} & \to & z_{t-1} & \to & z_t & \to & \cdots \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
\cdots & & x_{t-2} & & x_{t-1} & & x_t & & \cdots
\end{array}$$

$$\begin{aligned}
z_t &= \boldsymbol{A} z_{t-1} + w_t \\
x_t &= \boldsymbol{C} z_t + v_t \\
w_t &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{Q}) \\
v_t &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{R}) \\
z_0 &\sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)
\end{aligned}$$

Goal: Given $\boldsymbol{x} = (x_t)$, identify $\boldsymbol{A}, \boldsymbol{C}, \boldsymbol{Q}, \boldsymbol{R}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0$.

A More Pedestrian Approach to EM

Write $\boldsymbol{\theta} = (\boldsymbol{A}, \boldsymbol{C}, \boldsymbol{Q}, \boldsymbol{R}, \boldsymbol{\mu}_0, \Sigma_0)$. We want to maximize

$$\ell(\boldsymbol{\theta}) = \log \mathbb{P}(\boldsymbol{x} \mid \boldsymbol{\theta}),$$

but we don't know how to evaluate this function! We only know $\mathbb{P}(\boldsymbol{x}, \boldsymbol{z} \mid \boldsymbol{\theta})$ from the model.

A More Pedestrian Approach to EM

Write $\boldsymbol{\theta} = (\boldsymbol{A}, \boldsymbol{C}, \boldsymbol{Q}, \boldsymbol{R}, \boldsymbol{\mu}_0, \Sigma_0)$. We want to maximize

$$\ell(\boldsymbol{\theta}) = \log \mathbb{P}(\boldsymbol{x} \mid \boldsymbol{\theta}),$$

but we don't know how to evaluate this function! We only know $\mathbb{P}(\boldsymbol{x}, \boldsymbol{z} \mid \boldsymbol{\theta})$ from the model. But,

$$\cdots = \log \frac{\mathbb{P}(\boldsymbol{x}, \boldsymbol{z} \mid \boldsymbol{\theta})}{\mathbb{P}(\boldsymbol{z} \mid \boldsymbol{x}, \boldsymbol{\theta})} = \log \mathbb{P}(\boldsymbol{x}, \boldsymbol{z} \mid \boldsymbol{\theta}) - \log \mathbb{P}(\boldsymbol{z} \mid \boldsymbol{x}, \boldsymbol{\theta}).$$

If we knew what $\mathbb{P}(\boldsymbol{z} \mid \boldsymbol{x}, \boldsymbol{\theta})$ was, we could evaluate this. But we don't know what that is either!

A More Pedestrian Approach to EM

We want to maximize

$$\ell(\boldsymbol{\theta}) = \log \mathbb{P}(\boldsymbol{x}, \boldsymbol{z} \mid \boldsymbol{\theta}) - \log \mathbb{P}(\boldsymbol{z} \mid \boldsymbol{x}, \boldsymbol{\theta}).$$

All we know is how to find $\mathbb{P}(\boldsymbol{z} \mid \boldsymbol{x}, \boldsymbol{\theta})$ if we know $\boldsymbol{\theta}$ already (from Kalman smoothing). And only approximately!

A More Pedestrian Approach to EM

We want to maximize

$$\ell(\boldsymbol{\theta}) = \log \mathbb{P}(\boldsymbol{x}, \boldsymbol{z} \mid \boldsymbol{\theta}) - \log \mathbb{P}(\boldsymbol{z} \mid \boldsymbol{x}, \boldsymbol{\theta}).$$

All we know is how to find $\mathbb{P}(\boldsymbol{z} \mid \boldsymbol{x}, \boldsymbol{\theta})$ if we know $\boldsymbol{\theta}$ already (from Kalman smoothing). And only approximately! So maybe our only reliable operation is to take *expectations* with $\mathbb{P}(\boldsymbol{z} \mid \boldsymbol{x}, \boldsymbol{\theta})$.

A More Pedestrian Approach to EM

We want to maximize

$$\ell(\boldsymbol{\theta}) = \log \mathbb{P}(\boldsymbol{x}, \boldsymbol{z} \mid \boldsymbol{\theta}) - \log \mathbb{P}(\boldsymbol{z} \mid \boldsymbol{x}, \boldsymbol{\theta}).$$

All we know is how to find $\mathbb{P}(\boldsymbol{z} \mid \boldsymbol{x}, \boldsymbol{\theta})$ if we know $\boldsymbol{\theta}$ already (from Kalman smoothing). And only approximately! So maybe our only reliable operation is to take *expectations* with $\mathbb{P}(\boldsymbol{z} \mid \boldsymbol{x}, \boldsymbol{\theta})$.

Still, we can do...

$$\begin{aligned}\ell(\boldsymbol{\theta}) = \int & \log \mathbb{P}(\boldsymbol{x}, \boldsymbol{z} \mid \boldsymbol{\theta}) \cdot \mathbb{P}(\boldsymbol{z} \mid \boldsymbol{x}, \boldsymbol{\theta}) d\boldsymbol{z} \\ & - \int \log \mathbb{P}(\boldsymbol{z} \mid \boldsymbol{x}, \boldsymbol{\theta}) \cdot \mathbb{P}(\boldsymbol{z} \mid \boldsymbol{x}, \boldsymbol{\theta}) d\boldsymbol{z}.\end{aligned}$$

A More Pedestrian Approach to EM

We want to maximize

$$\begin{aligned}\ell(\boldsymbol{\theta}) = &\int \log \mathbb{P}(\boldsymbol{x}, \boldsymbol{z} \mid \boldsymbol{\theta}) \cdot \mathbb{P}(\boldsymbol{z} \mid \boldsymbol{x}, \boldsymbol{\theta}) d\boldsymbol{z} \\ &- \int \log \mathbb{P}(\boldsymbol{z} \mid \boldsymbol{x}, \boldsymbol{\theta}) \cdot \mathbb{P}(\boldsymbol{z} \mid \boldsymbol{x}, \boldsymbol{\theta}) d\boldsymbol{z}.\end{aligned}$$

Let's think of $\mathbb{P}(\boldsymbol{z} \mid \boldsymbol{x}, \boldsymbol{\theta}) = \mathbb{Q}(\boldsymbol{z})$ as a separate variable, not depending directly on $\boldsymbol{\theta}$.

A More Pedestrian Approach to EM

We want to maximize

$$\ell(\boldsymbol{\theta}) = \int \log \mathbb{P}(\boldsymbol{x}, \boldsymbol{z} \mid \boldsymbol{\theta}) \cdot \mathbb{P}(\boldsymbol{z} \mid \boldsymbol{x}, \boldsymbol{\theta}) d\boldsymbol{z}$$
$$- \int \log \mathbb{P}(\boldsymbol{z} \mid \boldsymbol{x}, \boldsymbol{\theta}) \cdot \mathbb{P}(\boldsymbol{z} \mid \boldsymbol{x}, \boldsymbol{\theta}) d\boldsymbol{z}.$$

Let's think of $\mathbb{P}(\boldsymbol{z} \mid \boldsymbol{x}, \boldsymbol{\theta}) = \mathbb{Q}(\boldsymbol{z})$ as a separate variable, not depending directly on $\boldsymbol{\theta}$. When we've maximized ℓ, we will have found two objects, $\boldsymbol{\theta}$ and $\mathbb{Q}$, that satisfy

$$\boldsymbol{\theta} \text{ is the maximizer of } \int \log \mathbb{P}(\boldsymbol{x}, \boldsymbol{z} \mid \boldsymbol{\theta}) \cdot \mathbb{Q}(\boldsymbol{z}) d\boldsymbol{z}$$
$$\mathbb{Q}(\boldsymbol{z}) = \mathbb{P}(\boldsymbol{z} \mid \boldsymbol{x}, \boldsymbol{\theta})$$

A More Pedestrian Approach to EM

Optimality conditions:

$$\boldsymbol{\theta} \text{ is the maximizer of } \int \log \mathbb{P}(\boldsymbol{x}, \boldsymbol{z} \mid \boldsymbol{\theta}) \cdot \mathbb{Q}(\boldsymbol{z}) d\boldsymbol{z}$$

$$\mathbb{Q}(\boldsymbol{z}) = \mathbb{P}(\boldsymbol{z} \mid \boldsymbol{x}, \boldsymbol{\theta})$$

And we know how to make either one of these true: either do the maximization (more on this in a second), or use Kalman smoothing (getting an approximate answer; more on this in a second also).

A More Pedestrian Approach to EM

Optimality conditions:

$$\boldsymbol{\theta} \text{ is the maximizer of } \int \log \mathbb{P}(\boldsymbol{x}, \boldsymbol{z} \mid \boldsymbol{\theta}) \cdot \mathbb{Q}(\boldsymbol{z}) d\boldsymbol{z}$$

$$\mathbb{Q}(\boldsymbol{z}) = \mathbb{P}(\boldsymbol{z} \mid \boldsymbol{x}, \boldsymbol{\theta})$$

And we know how to make either one of these true: either do the maximization (more on this in a second), or use Kalman smoothing (getting an approximate answer; more on this in a second also).

EM is then the obvious iterative algorithm to look for a fixed point of these conditions: just alternate making each one true.

There is extra structure in the LDS setup that lets us do “EM with a twist” even though we are limited in our ability to do some of these calculations.

Let’s take a look...

Nuances in the E Step

From $\boldsymbol{\theta}$, Kalman smoothing doesn't tell us how to compute $\mathbb{Q}(\boldsymbol{z}) = \mathbb{P}(\boldsymbol{z} \mid \boldsymbol{x}, \boldsymbol{\theta})$.

It only tells us how to compute some *partial statistics*.

Nuances in the E Step

From $\boldsymbol{\theta}$, Kalman smoothing doesn't tell us how to compute $\mathbb{Q}(\boldsymbol{z}) = \mathbb{P}(\boldsymbol{z} \mid \boldsymbol{x}, \boldsymbol{\theta})$.

It only tells us how to compute some *partial statistics*. First, we get *one point marginals*:

$$\mathbb{Q}(z_i) = \mathbb{P}(z_i \mid \boldsymbol{x}, \boldsymbol{\theta}).$$

Also, if you review the derivation, you'll see that we get (the covariances of) *two point marginals*, **but only for consecutive times**:

$$\mathbb{Q}(z_i, z_{i+1}) = \mathbb{P}((z_i, z_{i+1}) \mid \boldsymbol{x}, \boldsymbol{\theta}).$$

Nuances in the E Step

From $\boldsymbol{\theta}$, Kalman smoothing doesn't tell us how to compute $\mathbb{Q}(\boldsymbol{z}) = \mathbb{P}(\boldsymbol{z} \mid \boldsymbol{x}, \boldsymbol{\theta})$.

It only tells us how to compute some *partial statistics*. First, we get *one point marginals*:

$$\mathbb{Q}(z_i) = \mathbb{P}(z_i \mid \boldsymbol{x}, \boldsymbol{\theta}).$$

Also, if you review the derivation, you'll see that we get (the covariances of) *two point marginals*, **but only for consecutive times**:

$$\mathbb{Q}(z_i, z_{i+1}) = \mathbb{P}((z_i, z_{i+1}) \mid \boldsymbol{x}, \boldsymbol{\theta}).$$

Does that mean we're in trouble for the M step?

Nuances in the M Step

Let's recall the model:

$$\mathbb{P}(\boldsymbol{x}, \boldsymbol{z} \mid \boldsymbol{\theta}) = \mathbb{P}(z_0 \mid \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \cdot \prod_{i=0}^{n-1} \mathbb{P}(z_{i+1} \mid z_i, \boldsymbol{\theta}) \cdot \prod_{i=0}^{n} \mathbb{P}(x_i \mid z_i, \boldsymbol{\theta}).$$

(PS—writing down this formula from the graphical model is good practice!)

Nuances in the M Step

Let's recall the model:

$$\mathbb{P}(\boldsymbol{x}, \boldsymbol{z} \mid \boldsymbol{\theta}) = \mathbb{P}(z_0 \mid \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \cdot \prod_{i=0}^{n-1} \mathbb{P}(z_{i+1} \mid z_i, \boldsymbol{\theta}) \cdot \prod_{i=0}^{n} \mathbb{P}(x_i \mid z_i, \boldsymbol{\theta}).$$

(PS—writing down this formula from the graphical model is good practice!) From the evolution equations:

$$\begin{aligned} z_{i+1} &= \boldsymbol{A} z_i + w_i \\ x_i &= \boldsymbol{C} z_i + v_i \\ w_i &\sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}) \\ v_i &\sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}) \end{aligned}$$

Let's write down the log (on the board). What do you notice?

Summary: The Happy Coincidence of EM for LDS

1. EM is a general technique, but it requires us to be able to take the marginals of the latent variables in our model (the usual E step).

2. We can only do part of the job efficiently for LDS, getting only partial information on this marginal distribution.

3. However, that is exactly the part that we need for the calculations in the M step!

Coding Assignment

- Find `labs/lab4/lab4-student.ipynb` on the course Github. (It will be easier if you clone the whole repository.)
- This assignment just has two small bits of code to fill in, one gathering some matrices for the calculations, and one getting the values of transition and transition covariance matrices in the M step.
- Before you start, fill in your implementations of filter and smooth from last week's lab. If you don't have working implementations, look on with your neighbor for now, and I will post reference solutions later that you can use.