Algorithmic Trading: Expectation Maximization

Sebastian Jaimungal
University of Toronto

Jan, 2018

Expectation Maximization

EM Algorithm

- **Expectation Maximization** allows one to obtain **maximum likelihood estimates**(MLE) for models with **latent variables**
- It consists of two main steps:
  - **E-step** (expecatation)
    - estimate the latent variables given the observations
  - **M-step** (maximization)
    - maximize the likelihood given the estimated latent variables

EM Algorithm

- $\boldsymbol{X} = \{X_t : t = 1, \ldots, N\}$ are the r.v. corresponding to the observed data
- $\boldsymbol{x} = \{x_t : t = 1, \ldots, N\}$ are the **observed data**
- Θ is the set of **model parameters** which you aim to estimate
- $\boldsymbol{Z} = \{Z_t : t = 1, \ldots, N\}$ are the r.v. corresponding to the latent variables
- $z = \{z_t : t = 1, \ldots, N\}$ are the unobserved **latent variables**
- We aim to maximize the **log-likelihood**

$$\ell(\Theta) = \log \mathbb{P}(\boldsymbol{X} = \boldsymbol{x} \,|\, \Theta)$$

of the observed data

EM Algorithm

- Since $\boldsymbol{z}$ are unobserved, the likelihood consists of summing over all possible values

$$\mathbb{P}(\boldsymbol{X} = \boldsymbol{x} \,|\, \Theta) = \sum_{\boldsymbol{z}} \mathbb{P}(\boldsymbol{X} = \boldsymbol{x}\,,\; \boldsymbol{Z} = \boldsymbol{z} \,|\, \Theta)$$

- Instead, EM seeks to construct a **sequence of improvements**

$$\begin{aligned}
\ell(\Theta) &= \log \sum_{\boldsymbol{z}} \mathbb{P}(\boldsymbol{X} = \boldsymbol{x}\,,\; \boldsymbol{Z} = \boldsymbol{z} \,|\, \Theta) \\
&= \log \sum_{\boldsymbol{z}} \mathbb{Q}(\boldsymbol{Z} = \boldsymbol{z} \mid \boldsymbol{X} = \boldsymbol{x}) \, \frac{\mathbb{P}(\boldsymbol{X} = \boldsymbol{x}\,,\; \boldsymbol{Z} = \boldsymbol{z} \,|\, \Theta)}{\mathbb{Q}(\boldsymbol{Z} = \boldsymbol{z} \mid \boldsymbol{X} = \boldsymbol{x})} \\
&= \log \mathbb{E}^{\mathbb{Q}}[\, \Psi_{\mathbb{Q}}(\boldsymbol{X} = \boldsymbol{x},\, \boldsymbol{Z}; \Theta) \,]
\end{aligned}$$

here, $\mathbb{Q}$ is **any** probability distribution over the latent variables, and the r.v.

$$\Psi_{\mathbb{Q}}(\boldsymbol{X} = \boldsymbol{x},\, \boldsymbol{Z} = \cdot\,; \Theta) = \frac{\mathbb{P}(\boldsymbol{X} = \boldsymbol{x}\,,\; \boldsymbol{Z} = \cdot \,|\, \Theta)}{\mathbb{Q}(\boldsymbol{Z} = \cdot \mid \boldsymbol{X} = \boldsymbol{x})}$$

EM Algorithm

- **Jensen's inequality** gives

$$\ell(\Theta) \geq \mathbb{E}^{\mathbb{Q}}[\,\log \Psi_{\mathbb{Q}}(\boldsymbol{X} = \boldsymbol{x},\ \boldsymbol{Z}; \Theta)\,]$$

- The lower bound is saturated if $\Psi_{\mathbb{Q}}(\boldsymbol{X} = \boldsymbol{x},\ \boldsymbol{Z}; \Theta) = const.$, i.e., when

$$\mathbb{Q}_{\Theta}(\cdot) = \frac{\mathbb{P}(\boldsymbol{X} = \boldsymbol{x}\ ,\ \boldsymbol{Z} = \cdot \,|\, \Theta)}{\sum_{\boldsymbol{z}} \mathbb{P}(\boldsymbol{X} = \boldsymbol{x}\ ,\ \boldsymbol{Z} = \boldsymbol{z} \,|\, \Theta)} = \mathbb{P}(\boldsymbol{Z} = \cdot \,|\, \boldsymbol{X} = \boldsymbol{x};\ \Theta)$$

- Hence,

$$\ell(\Theta) = \mathbb{E}^{\mathbb{Q}_{\Theta}}[\,\log \Psi_{\mathbb{Q}_{\Theta}}(\boldsymbol{X} = \boldsymbol{x}; \Theta)\,]$$

EM Algorithm

- Next, suppose we have a current estimate Θ_k for the MLE, then consider the new function

$$\begin{aligned}\bar{\ell}(\Theta) &= \mathbb{E}^{\mathbb{Q}_{\Theta_k}}[\, \log \Psi_{\mathbb{Q}_{\Theta_k}}(\boldsymbol{X} = \boldsymbol{x}; \Theta)\,] \\ &= \mathbb{E}^{\mathbb{Q}_{\Theta_k}}\left[\, \log \frac{\mathbb{P}(\boldsymbol{X} = \boldsymbol{x}\,,\ \boldsymbol{Z} \,|\, \Theta)}{\mathbb{Q}_{\Theta_k}(\boldsymbol{Z} \mid \boldsymbol{X} = \boldsymbol{x})}\,\right]\end{aligned}$$

- Consider a new estimate Θ_{k+1} given by

$$\Theta_{k+1} = \arg\max_{\Theta} \bar{\ell}(\Theta)$$

- Would like to show that $\ell(\Theta_{k+1}) \geq \ell(\Theta_k)$, i.e., that this estimate improves the log-likelihood

EM Algorithm

- To this end, we have for any $\mathbb{Q}$

$$\ell(\Theta_{k+1}) \geq \mathbb{E}^{\mathbb{Q}}[\,\log \Psi_{\mathbb{Q}}(\boldsymbol{X} = \boldsymbol{x}; \Theta_{k+1})\,]$$

so certainly it is true for $\mathbb{Q} = \mathbb{Q}_k$, and hence

$$\begin{aligned}
\ell(\Theta_{k+1}) &\geq \mathbb{E}^{\mathbb{Q}_k}[\,\log \Psi_{\mathbb{Q}_k}(\boldsymbol{X} = \boldsymbol{x}; \Theta_{k+1})\,] \\
&= \bar{\ell}(\Theta_{k+1}) \\
&\geq \bar{\ell}(\Theta_k) \qquad \because \Theta_{k+1} \text{ maximizes } \bar{\ell} \\
&= \ell(\Theta_k)
\end{aligned}$$