Geometry of EM and em algorithms

Salem Said

Nonlinear Signal Processing Lab
University of Melbourne

Differential and Information Geometry 2011

What is EM

Numerical method for maximum likelihood, without derivatives

Objective

$$\ell(\eta) = \log[p_\eta(y)] \qquad \eta^* = argmax_\eta \ell(\eta)$$

Idea Introduce hidden variable x where $x = f(y)$

$$\ell(\eta, \eta^0) = \mathbb{E}_{\eta^0}\{\log[p_\eta(x)] | y\} \qquad \eta^1 = argmax_\eta \ell(\eta, \eta^0) \qquad (1)$$

E and M steps η^0 initial guess

E $\Rightarrow$ compute $\ell(\eta|\eta^0)$ M $\Rightarrow$ obtain η^1

Monotonicity

$$\ell(\eta, \eta^0) \geq \ell(\eta^0, \eta^0) \Rightarrow \ell(\eta) \geq \ell(\eta^0)$$

What is EM

Numerical method for maximum likelihood, without derivatives

Objective

$$\ell(\eta) = \log[p_\eta(y)] \qquad \eta^* = argmax_\eta \ell(\eta)$$

Idea Introduce hidden variable x where $x = f(y)$

$$\ell(\eta, \eta^0) = \mathbb{E}_{\eta^0}\{\log[p_\eta(x)]|y\} \qquad \eta^1 = argmax_\eta \ell(\eta, \eta^0) \qquad (1)$$

E and M steps η^0 initial guess

E $\Rightarrow$ compute $\ell(\eta|\eta^0)$ M $\Rightarrow$ obtain η^1

Monotonicity

$$\ell(\eta, \eta^0) \geq \ell(\eta^0, \eta^0) \Rightarrow \ell(\eta) \geq \ell(\eta^0)$$

What is EM

Numerical method for maximum likelihood, without derivatives

Objective

$$\ell(\eta) = \log[p_\eta(y)] \quad \eta^* = argmax_\eta \ell(\eta)$$

Idea Introduce hidden variable x where $x = f(y)$

$$\ell(\eta, \eta^0) = \mathbb{E}_{\eta^0}\{\log[p_\eta(x)]|y\} \quad \eta^1 = argmax_\eta \ell(\eta, \eta^0) \tag{1}$$

E and M steps η^0 initial guess

E $\Rightarrow$ compute $\ell(\eta|\eta^0)$ M $\Rightarrow$ obtain η^1

Monotonicity

$$\ell(\eta, \eta^0) \geq \ell(\eta^0, \eta^0) \Rightarrow \ell(\eta) \geq \ell(\eta^0)$$

What is EM

Numerical method for maximum likelihood, without derivatives

Objective

$$\ell(\eta) = \log[p_\eta(y)] \qquad \eta^* = argmax_\eta \ell(\eta)$$

Idea Introduce hidden variable x where $x = f(y)$

$$\ell(\eta, \eta^0) = \mathbb{E}_{\eta^0}\{\log[p_\eta(x)]|y\} \qquad \eta^1 = argmax_\eta \ell(\eta, \eta^0) \qquad (1)$$

E and M steps η^0 initial guess

E $\Rightarrow$ compute $\ell(\eta|\eta^0)$ M $\Rightarrow$ obtain η^1

Monotonicity

$$\ell(\eta, \eta^0) \geq \ell(\eta^0, \eta^0) \Rightarrow \ell(\eta) \geq \ell(\eta^0)$$

EM in an exponential family

$$p(t|\eta) = \exp\left[\sum_{k=1}^{s} \eta_k t_k - \psi(\eta)\right]$$

E and M steps

$$\theta(u^0) = \mathbb{E}_{u^0}\{t|x_v\} \qquad D(\theta(u^0)|\eta(u^1)) = \min_u D(\theta(u^0)|\eta(u^1)) \quad (2)$$

Partial observation $t = t(x_v, x_h)$

$$D = \{\theta = t(x_v|x_h)|x_h \in X\} \qquad x = (x_v, x_h)$$

Curved Model

$$M = \{\eta = \eta(u)|u \in O\} \qquad O \subset \mathbb{R}^{s'}$$

EM in an exponential family

$$p(t|\eta) = \exp\left[\sum_{k=1}^{s} \eta_k t_k - \psi(\eta)\right]$$

E and M steps

$$\theta(u^0) = \mathbb{E}_{u^0}\{t|x_v\} \qquad D(\theta(u^0)|\eta(u^1)) = \min_u D(\theta(u^0)|\eta(u^1)) \quad (2)$$

Partial observation $t = t(x_v, x_h)$

$$D = \{\theta = t(x_v|x_h)|x_h \in X\} \qquad x = (x_v, x_h)$$

Curved Model

$$M = \{\eta = \eta(u)|u \in O\} \qquad O \subset \mathbb{R}^{s'}$$

EM in an exponential family

$$p(t|\eta) = \exp\left[\sum_{k=1}^{s} \eta_k t_k - \psi(\eta)\right]$$

E and M steps

$$\theta(u^0) = \mathbb{E}_{u^0}\{t|x_v\} \qquad D(\theta(u^0)|\eta(u^1)) = \min_u D(\theta(u^0)|\eta(u^1)) \quad (2)$$

Partial observation $t = t(x_v, x_h)$

$$D = \{\theta = t(x_v|x_h)|x_h \in X\} \qquad x = (x_v, x_h)$$

Curved Model

$$M = \{\eta = \eta(u)|u \in O\} \qquad O \subset \mathbb{R}^{s'}$$

Dual coordinates

Legendre pair

$$\phi(\theta) = \max_{\eta}\{(\theta,\eta) - \psi(\eta) \qquad \psi(\eta) = \max_{\theta}\{(\eta,\theta) - \phi(\theta)$$

Dual coordinates (e and m coordinates)

$$\theta(\eta) = \nabla\psi(\eta) \qquad \eta(\theta) = \nabla\phi(\theta)$$

Fisher information

$$g_{kl}(\eta) = \partial_k \partial_l \psi(\eta) \qquad g^{kl}(\theta) = \partial_k \partial_l \phi(\theta)$$

e and m geodesics: straight lines in η and θ

Observed manifold

$$D = \{\theta = t(x_v|x_h)|x_h \in X\} \quad x = (x_v, x_h)$$

m projection $D \to M$

$$\theta^* \in D \mapsto \eta(u^*) \quad D(\theta^*|\eta(u^*)) = \min_M D(\theta^*|\eta(u)) \qquad (3)$$

e projection $M \to D$

$$\eta(u) \in M \mapsto \theta(u) \quad D(theta(u)|\eta(u)) = \min_D D(\theta|\eta(u)) \qquad (4)$$

- m projection: m geodesic from θ orthogonal to M
- e projection: e geodesic from η orthogonal to D

Observed manifold

$$D = \{\theta = t(x_v|x_h)|x_h \in X\} \quad x = (x_v, x_h)$$

m projection $D \to M$

$$\theta^* \in D \mapsto \eta(u^*) \quad D(\theta^*|\eta(u^*)) = \min_M D(\theta^*|\eta(u)) \tag{3}$$

e projection $M \to D$

$$\eta(u) \in M \mapsto \theta(u) \quad D(theta(u)|\eta(u)) = \min_D D(\theta|\eta(u)) \tag{4}$$

- m projection: m geodesic from θ orthogonal to M
- e projection: e geodesic from η orthogonal to D

Observed manifold

$$D = \{\theta = t(x_v|x_h)|x_h \in X\} \quad x = (x_v, x_h)$$

m projection $D \rightarrow M$

$$\theta^* \in D \mapsto \eta(u^*) \quad D(\theta^*|\eta(u^*)) = \min_M D(\theta^*|\eta(u)) \tag{3}$$

e projection $M \rightarrow D$

$$\eta(u) \in M \mapsto \theta(u) \quad D(theta(u)|\eta(u)) = \min_D D(\theta|\eta(u)) \tag{4}$$

- m projection: m geodesic from θ orthogonal to M
- e projection: e geodesic from η orthogonal to D

em algorithm

Objective

$$D(\theta^*|\eta(u^*)) = \min_{\theta \in D, u \in M}(\theta|\eta(u))$$

e and m steps Iterate e and m projections

- m projection: $\theta^1 \mapsto \eta(u^1) \in M$
- e projection: $\eta(u^1) \mapsto \theta(u^1) \in D$

Property of e projection (projection η^*, θ^*)

Assume $D = (t_v, \theta_h)$ and write $\eta = (\eta_v, \eta_h)$

$$\theta^*_v = \theta_v \qquad \eta^*_h = \eta_h \tag{5}$$

$$\mathbb{E}_\eta\{t_h|t_v\} = \mathbb{E}_{\eta^*}\{t_h|t_v\} \tag{6}$$

em algorithm

Objective

$$D(\theta^*|\eta(u^*)) = \min_{\theta\in D, u\in M} (\theta|\eta(u))$$

e and m steps Iterate e and m projections

- m projection: $\theta^1 \mapsto \eta(u^1) \in M$
- e projection: $\eta(u^1) \mapsto \theta(u^1) \in D$

Property of e projection (projection η^*, θ^*)

Assume $D = (t_v, \theta_h)$ and write $\eta = (\eta_v, \eta_h)$

$$\theta_v^* = \theta_v \qquad \eta_h^* = \eta_h \tag{5}$$

$$\mathbb{E}_\eta\{t_h|t_v\} = \mathbb{E}_{\eta^*}\{t_h|t_v\} \tag{6}$$

em algorithm

Objective

$$D(\theta^*|\eta(u^*)) = \min_{\theta \in D, u \in M} (\theta|\eta(u))$$

e and m steps Iterate e and m projections

- m projection: $\theta^1 \mapsto \eta(u^1) \in M$
- e projection: $\eta(u^1) \mapsto \theta(u^1) \in D$

Property of e projection (projection η^*, θ^*)

Assume $D = (t_v, \theta_h)$ and write $\eta = (\eta_v, \eta_h)$

$$\theta_v^* = \theta_v \quad \eta_h^* = \eta_h \tag{5}$$

$$\mathbb{E}_\eta\{t_h|t_v\} = \mathbb{E}_{\eta^*}\{t_h|t_v\} \tag{6}$$

Geometry of EM

m and M steps are the same, when are e and E steps the same?

$$F(\theta_v, \theta_h) = (\theta_v, \mathbb{E}_\eta\{t_h|\theta_v\}) \in D$$

E step

$$D(F(\theta(u^1))|\eta(u^1)) = \min_D D(F(\theta)|\eta(u^1)) \quad (7)$$

e step

$$D(\theta(u^1)|\eta(u^1)) = \min_D D(\theta|\eta(u^1)) \quad (8)$$

- E step is the same as e step *iff* F is the identity on D
- This is equivalent to

$$\mathbb{E}_\eta\{t_h|\theta_v\}) = A\theta_v + b$$

Geometry of EM

m and M steps are the same, when are e and E steps the same?

$$F(\theta_v, \theta_h) = (\theta_v, \mathbb{E}_\eta\{t_h|\theta_v\}) \in D$$

E step

$$D(F(\theta(u^1))|\eta(u^1)) = \min_D D(F(\theta)|\eta(u^1)) \quad (7)$$

e step

$$D(\theta(u^1)|\eta(u^1)) = \min_D D(\theta|\eta(u^1)) \quad (8)$$

- E step is the same as e step *iff* F is the identity on D
- This is equivalent to

$$\mathbb{E}_\eta\{t_h|\theta_v\}) = A\theta_v + b$$

Geometry of EM

m and M steps are the same, when are e and E steps the same?

$$F(\theta_v, \theta_h) = (\theta_v, \mathbb{E}_\eta\{t_h|\theta_v\}) \in D$$

E step

$$D(F(\theta(u^1))|\eta(u^1)) = \min_D D(F(\theta)|\eta(u^1)) \tag{7}$$

e step

$$D(\theta(u^1)|\eta(u^1)) = \min_D D(\theta|\eta(u^1)) \tag{8}$$

- E step is the same as e step *iff* F is the identity on D
- This is equivalent to

$$\mathbb{E}_\eta\{t_h|\theta_v\}) = A\theta_v + b$$

Geometry of EM

m and M steps are the same, when are e and E steps the same?

$$F(\theta_v, \theta_h) = (\theta_v, \mathbb{E}_\eta\{t_h|\theta_v\}) \in D$$

E step

$$D(F(\theta(u^1))|\eta(u^1)) = \min_D D(F(\theta)|\eta(u^1)) \tag{7}$$

e step

$$D(\theta(u^1)|\eta(u^1)) = \min_D D(\theta|\eta(u^1)) \tag{8}$$

- E step is the same as e step *iff* F is the identity on D
- This is equivalent to

$$\mathbb{E}_\eta\{t_h|\theta_v\}) = A\theta_v + b$$