2/5/07

Recall Alt Min & EM

(1) Alt Min

$$\max_{\theta, X} p_\theta(Y, X)$$ ← applicable to any split of parameters $\theta = [\theta_1, \theta_2]$

If this cannot be computed easily (efficiently), use alternating minimization. Usually at least one (sometimes both) steps of Alt Min can be solved in closed form.

Idea: (1) Hold θ^k fixed, find

$$X^{k+1} = \arg\max_X p_{\theta_k}(Y, X)$$

(2) Hold X^{k+1} fixed, find

$$\theta^{k+1} = \arg\max_\theta p_\theta(Y, X^{k+1})$$

(3) $k \leftarrow k+1$, go to (1)

Stop if $\|\theta^k - \theta^{k-1}\|$ below a tolerance

<same as: $\|p_{\theta^k}(Y, X^k) - p_{\theta^{k-1}}(Y, X^{k-1})\| <$ tolerance>

(2) EM: Goal is to find $\max_\theta p_\theta(Y)$.

Y : data.

This maximization is difficult but can be simplified by conditioning on hidden data (X) & doing "lower bound optimization"

$$\max_\theta p_\theta(Y) = \max_\theta \sum_X p_\theta(Y, X) = \max_\theta \sum_X f(X) \frac{p_\theta(Y, X)}{f(X)}$$

$$\max_{\underline{\theta}} p_{\underline{\theta}}(\underline{y}) = \max_{\underline{\theta}} \sum_{\underline{x}} f(\underline{x}) \frac{p_\theta(\underline{y}, \underline{x})}{f(\underline{x})} \geq \max_{\underline{\theta}} \sum_{\underline{x}} f(\underline{x}) \log \frac{p_\theta(\underline{y}, \underline{x})}{f(\underline{x})}$$

$f(\underline{x})$: any PMF on $\underline{x}$.

holds because $f(\underline{x})$ is a PMF
< log-sum ineq or Jensen's ineq >.

$$\max_{\underline{\theta}} p_{\underline{\theta}}(\underline{y}) \geq \max_{\underline{\theta}} \sum_{\underline{x}} f(\underline{x}) \log \frac{p_{\underline{\theta}}(\underline{y}, \underline{x})}{f(\underline{x})}$$

$$\geq \max_{\substack{\underline{\theta},\, f(\underline{x}), \\ \text{s.t. } \sum_{\underline{x}} f(\underline{x}) = 1}} \sum_{\underline{x}} f(\underline{x}) \log \frac{p_\theta(\underline{y}, \underline{x})}{f(\underline{x})}$$

$\longrightarrow$ Optimize the lower bound at each step using Alternating Minimiz^n on $[\underline{\theta}, f(\underline{x})]$:
Result is the EM algorithm.

EM steps

① E step: $q(\underline{\theta}, \underline{\theta}^k) = E_{p_{\underline{\theta}^k}(\underline{x}|\underline{y})} [\log p_{\underline{\theta}}(\underline{y}, \underline{x})]$

② M step: $\underline{\theta}^{k+1} = \arg\max_{\underline{\theta}} q(\underline{\theta}, \underline{\theta}^k)$

③ $k \leftarrow k+1$, go to step ①

Stop if $\|\underline{\theta}^k - \underline{\theta}^{k-1}\|$ less than tolerance.
< also same as $\|q(\underline{\theta}^k, \underline{\theta}^{k-1}) - q(\theta^{k-1}, \theta^{k-2})\|$ < tolerance

Applications to Segmentation

→ **Data** $\underline{Y}$ = pixel intensities, or color or "features"

Hidden data $\underline{X}$ = Class labels.

parameters $\underline{\theta}$ = Gaussian mixture model parameters

$p_c, \mu_c, \Sigma_c \quad c = 1, 2, \ldots K.$

< Each object "c" generated from a Gaussian with mean μ_c, covariance Σ_c & prior probability p_c. >.

$\underline{Y}$ can also be texture feature, ~~such~~ or Gabor filters' outputs or DCT coefficients of 8x8 blocks, or motion features (e.g. optical flow or affine motion model)

Other applications of Alt. Max

Cost function may not be $p_{\underline{\theta}}(\underline{Y}, \underline{X})$: can be any $f(\underline{\theta}_1, \underline{\theta}_2)$.

~~each~~ e.g. a) Solve for correspondence & register

b) segment & register where segmenting by finding the "optimal" contour

c) Clustering for object recognition

(d) Vector Quantization, information theory, ---

Connection b/w EM & Alt-Min ~~for~~

EM : ~~find~~

$$\underline{\theta}^{k+1} = \arg\max_{\underline{\theta}} E_{p_{\theta^k}(x|y)}\left[\log p_{\underline{\theta}}(\underline{Y}, \underline{X})\right]$$

If $p_{\theta^k}(\underline{x}|\underline{y})$ is very peaky, i.e.

$$p_{\underline{\theta}^k}(\underline{x}|\underline{y}) \approx \delta(\underline{x} - \underline{x}^k), \quad \underline{x}^k = \arg\max_{\underline{x}} p_{\underline{\theta}^k}(\underline{x}|\underline{y}),$$

δ: Kronecker or Dirac delta fn
($\underline{x}$ discrete or continuous)

< Note $\underline{x}^k = \arg\max_{\underline{x}} p_{\underline{\theta}^k}(\underline{x}|\underline{Y}) = \arg\max_{\underline{x}} p_{\underline{\theta}^k}(\underline{Y}, \underline{X})$ >

then,

$$\underline{\theta}^{k+1} = \arg\max_{\underline{\theta}} \log p_{\underline{\theta}}(\underline{Y}, \underline{x}^k)$$

where $\underline{X}^k = \arg\max_{\underline{X}} p_{\theta^k}(\underline{Y}, \underline{X})$

Alt Max. of $p_\theta(\underline{Y}, \underline{x})$

Thus whenever $p_{\theta^k}(\underline{x}|\underline{y})$ is peaky (clear clusters exist), EM $\simeq$ Alt Max over $p_{\underline{\theta}}(\underline{Y}, \underline{X})$