ECE 544NA
9/5/2013

TODAY: PARAMETRIC II

I. MAP / BAYESIAN ESTIMATION
II. MIXTURE DENSITIES
III. VARIATIONAL BAYES & EM

ML ESTIMATION: $\hat{\Theta} = \arg\max p(\underline{X} \mid \Theta)$

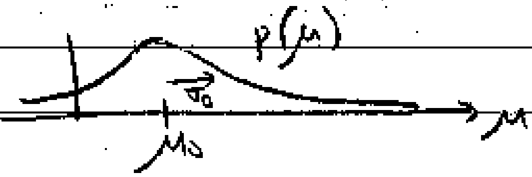
MAP ESTIMATION $\hat{\Theta} = \arg\max p(\Theta \mid \underline{X})$

$$= \arg\max \frac{p(\underline{X} \mid \Theta)\, p(\Theta)}{p(\underline{X})}$$

$$= \arg\max p(\underline{X} \mid \Theta)\, p(\Theta)$$

$$= \arg\max \left(\ln p(\Theta) + \sum_{n=1}^{N} \ln p(\vec{x}_n \mid \Theta) \right)$$

WHAT IS $p(\Theta)$?



① CENTROID = VALUE OF Θ YOU CONSIDER MOST PROBABLE

② SPREAD $\propto \frac{1}{\text{CONFIDENCE}}$ OF YOUR A PRIORI ESTIMATE

③ CONJUGATE PRIOR $\equiv p(\Theta)$ CHOSEN SO THAT $\nabla_{\Theta} \ln p(\Theta)$ HAS THE SAME DEPENDENCE ON Θ AS $\nabla_{\Theta} \ln p(\vec{x}_n \mid \Theta)$

EXAMPLES

① GAUSSIAN MEAN

$\boxed{\text{CONJUGATE PRIOR} = \text{GAUSSIAN}}$

$$p(\mu) = \frac{1}{(2\pi)^{D/2}|\Sigma_0|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mu-\mu_0)^T \Sigma_0^{-1}(\mu-\mu_0)}$$

$\mu_0, \Sigma_0 =$ "HYPERPARAMETERS"

$$\nabla_\mu \left(\ln p(\mu) + \sum_{n=1}^{N} \ln p(\vec{x}_n|\mu) \right)$$

$$= \Sigma_0^{-1}(\vec{\mu}_0 - \vec{\mu}) + \Sigma^{-1} \sum_{n=1}^{N} (\vec{x}_n - \vec{\mu})$$

$$= 0 \quad \text{IF} \quad \boxed{\hat{\mu} = (\tau + N)^{-1}\left(\tau \mu_0 + \sum_{n=1}^{N} \vec{x}_n\right) \quad \tau = \Sigma\Sigma_0^{-1}}$$

② GAUSSIAN COVARIANCE

CONJUGATE PRIOR = WISHART DISTRIBUTION

$$p(\Sigma) \propto \frac{|R_0|^{\alpha/2}}{|\Sigma|^{\alpha/2}} e^{-\frac{\alpha}{2}\operatorname{trace}(\Sigma^{-1}R_0)}$$

HYPERPARAMETERS α, R_0

$$p(\vec{x}_n|\Sigma) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}\operatorname{trace}\left((x_n-\mu)^T\Sigma^{-1}(x_n-\mu)\right)}$$

$$= \frac{1}{(2\pi)^{D/2}|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}\operatorname{trace}\left(\Sigma^{-1}(x_n-\mu)(x_n-\mu)^T\right)}$$

TRACE IDENTITY: $\operatorname{trace}(AB) = \operatorname{trace}(BA)$

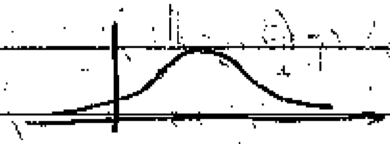
$$\therefore \quad \hat{\Sigma} = \frac{1}{\alpha+N}\left(\alpha R_0 + \sum_{n=1}^{N}(\vec{x}_n - \vec{\mu})(\vec{x}_n - \vec{\mu})^T\right)$$

CONFIDENCE

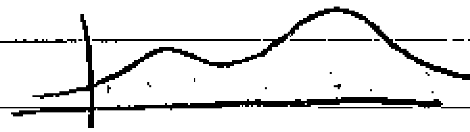
PRIOR BELIEF

II. MIXTURE DENSITIES

GAUSSIAN, LAPLACIAN, ... → UNIMODAL!



HOW TO REPRESENT

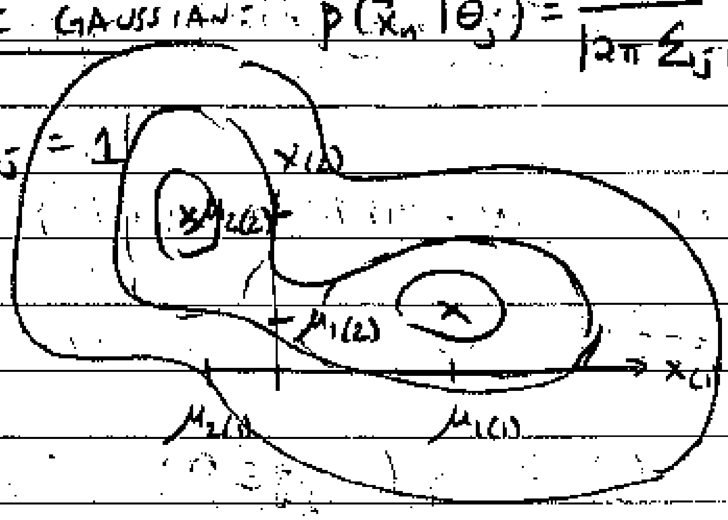


?

ANSWER: $p(\vec{x}_n \mid \theta) = \sum_{j=1}^{J} c_j \, p(\vec{x}_n \mid \theta_j)$

MIXTURE GAUSSIAN: $p(\vec{x}_n \mid \theta_j) = \frac{1}{|2\pi \Sigma_j|^{\frac{1}{2}}} e^{-\frac{1}{2}(\vec{x}_n - \vec{\mu}_j)^T \Sigma_j^{-1} (\vec{x}_n - \vec{\mu}_j)}$

$\sum_{j=1}^{J} c_j = 1$



ESTIMATION $\quad G(\underline{X}, \theta) =$

$$\ln p(\underline{X} \mid \theta) = \sum_{n=1}^{N} \ln p(x_n \mid \theta)$$

$$= \sum_{n=1}^{N} \ln \left(\sum_{j=1}^{J} c_j \, p(x_n \mid \theta_j) \right)$$

$\frac{\partial G(\underline{X}, \theta)}{\partial \theta} = 0$ ALMOST GIVES A SOLUTION, BUT NOT QUITE

III. VARIATIONAL BAYES

$$\sum_{j=1}^{J} q(j) = 1$$

TRICK: INTRODUCE $q(j)$ = PSEUDO-DISTRIBUTION

$$\ln p(\underline{X}|\theta) = \sum_{n=1}^{N} \ln \left(\sum_{j=1}^{J} q(j) \frac{p(\vec{x}_n, j|\theta)}{q(j)} \right)$$

JENSEN'S INEQUALITY: SINCE $\sum_{j=1}^{J} q(j) = 1$,

$$\ln \left(\sum_{j=1}^{J} q(j) f_j \right) \geq \sum_{j=1}^{J} q(j) \ln f_j$$

SO

VARIATIONAL BAYES

$$\boxed{\ln p(\underline{X}|\theta) \geq \sum_{j=1}^{J} \sum_{n=1}^{N} q(j) \ln \left(\frac{c_j\, p(\vec{x}_n|\theta, j)}{q(j)} \right)}$$

EXPECTATION MAXIMIZATION (EM)

1. $\boxed{\text{E-STEP}}$ FIND $q(j)$ TO MAXIMIZE RHS

$\Rightarrow$ (a) $q(j) \propto p(\vec{x}_n, j|\theta)$

$\boxed{\text{BUT}}$

(b) $\sum_{j=1}^{J} q(j) = 1$

$$(a)+(b) \Rightarrow \boxed{q(j) = p(j|\vec{x}_n, \theta) = \frac{p(\vec{x}_n, j|\theta)}{\sum_{j=1}^{J} p(\vec{x}_n, j|\theta)}}$$

2. $\boxed{\text{M-STEP}}$ FIND $\hat{\theta}$ TO MAXIMIZE RHS

$$\Rightarrow \boxed{\hat{\theta} = \operatorname{argmax} \sum_{j=1}^{J} \sum_{n=1}^{N} q(j) \ln p(\vec{x}_n, j|\theta)}$$

EQUIVALENTLY

① E-STEP COMPUTE $p(j \mid \vec{x}_n, \theta)$

② M-STEP FIND

$$\hat{\theta} = \arg\max Q(\theta, \hat{\theta})$$

$$= \arg\max \sum_{n=1}^{N} E_j\left[\ln p(\vec{x}_n, j \mid \hat{\theta}) \,\middle|\, \theta\right]$$

GAUSSIAN MIXTURE MODEL

① E-STEP $\mathcal{N}(\vec{x}_n \mid \vec{\mu}_j, \Sigma_j) = \frac{1}{|2\pi\Sigma_j|^{\frac{1}{2}}} e^{-\frac{1}{2}(\vec{x}_n - \vec{\mu}_j)^T \Sigma_j^{-1} (\vec{x}_n - \vec{\mu}_j)}$

$$\gamma_n(j) = \frac{c_j \mathcal{N}(\vec{x}_n \mid \vec{\mu}_j, \Sigma_j)}{\sum_{k=1}^{J} c_k \mathcal{N}(\vec{x}_n \mid \vec{\mu}_k, \Sigma_k)} = p(j \mid \vec{x}_n, \theta)$$

② M-STEP

$$\hat{c}_j = \frac{1}{N} \sum_{n=1}^{N} \gamma_n(j)$$

$$\hat{\mu}_j = \frac{\sum_{n=1}^{N} \gamma_n(j)\, x_n}{\sum_{n=1}^{N} \gamma_n(j)}$$

$$\hat{\Sigma}_j = \frac{\sum_{n=1}^{N} \gamma_n(j)\, (\vec{x}_n - \vec{\mu}_j)(\vec{x}_n - \vec{\mu}_j)^T}{\sum_{n=1}^{N} \gamma_n(j)}$$