GAMMA DENSITY

$$f(x;\alpha,\beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \qquad x \geq 0$$

Suppose $X_i \sim \exp(\lambda)$, $\lambda \sim \text{Gamma}(\alpha, b)$

$$f(x;\lambda) = \lambda e^{-\lambda x} \qquad x \geq 0$$

Find posterior

$$X_1, \ldots, X_n \text{ IID } \exp(\lambda)$$

$$f(x_1, \ldots, x_n; \lambda) = \lambda^n e^{\lambda \sum x_i}$$

$$f(\lambda \mid x_1, \ldots, x_n) = \frac{f(x_1, \ldots, x_n \mid \lambda) f(\lambda)}{\int_0^\infty f(x_1, \ldots, x_n \mid \lambda) f(\lambda) d\lambda}$$

$$\propto \lambda^n e^{-\lambda \sum x_i} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$$

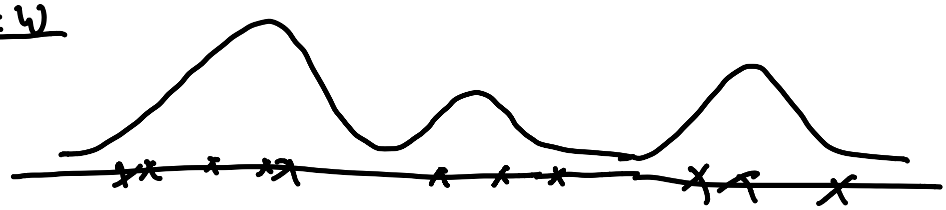
$$\propto \underbrace{\lambda^{n+\alpha-1} e^{-(\sum x_i + \beta)\lambda}}_{\text{Gamma}(n+\alpha,\ \sum x_i + \beta)}$$

$$= \frac{(\sum x_i + \beta)^{n+\alpha}}{\Gamma(n+\alpha)} \lambda^{n+\alpha-1} e^{-(\sum x_i + \beta)\lambda}$$

$$\mathbb{E}[\lambda \mid x_1, \ldots, x_n] = \frac{n+\alpha}{\sum x_i + \beta}$$

$$= \frac{1 + \alpha/n}{\overline{X} + \beta/n}$$

REVIEW



MIXTURE MODELS

$$f\left(x ; \begin{bmatrix}\mu_1\\ \mu_2\\ \mu_3\end{bmatrix}, \begin{bmatrix}\sigma_1^2\\ \sigma_2^2\\ \sigma_3^2\end{bmatrix}, \pi_1, \pi_2, \pi_3\right)$$

θ = all parameters

K NORMALS (1 DIMENSIONAL)

$\mu_j, \sigma_j^2 \quad j = 1, 2, \ldots, K$

π is a distribution on $\{1, 2, \ldots, K\}$

$$f(x ; \theta) = \sum_{j=1}^{K} \pi_j f(x ; \mu_j, \Sigma_j)$$

GOAL: FIND THE MLE

IE VALUE θ WHICH MAKE $L(\theta)$ as large as possible

[SEEKING LOCAL MAXIMUM IN THIS CASE]

$$L(\theta) = L(\theta ; \underset{\sim}{x}) = f(\underset{\sim}{x} ; \theta)$$

IE IF YOU HAVE IID $X_1, \ldots, X_n$

EACH has DENSITY AS ABOVE

$$f(x_1, \ldots, x_n ; \theta) = \prod_{i=1}^{n} f(x_i ; \theta) = L(\theta)$$

MLE IS THE VALUE θ WHICH MAXIMIZES $L(\theta)$

NO ANALYTIC SOLUTION

If $K=1$:

$$\mu = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \sigma^2 = \frac{1}{n}\sum_i (X_i - \bar{X})^2$$

If $K=1$, X_i is d-dimensional Normal

specified by two things

μ = vector of means d-dimensional vector

$$X_i = (X_i^{(1)}, \ldots, X_i^{(d)})$$

$$\Rightarrow \mu = (\mathbb{E}X_i^{(1)}, \ldots, \mathbb{E}X_i^{(d)})$$

$\underset{d \times d}{\Sigma}$ matrix $\Sigma_{jl} = \mathrm{Cov}(X_i^{(j)}, X_i^{(l)})$

$$f(x) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2} \underset{1 \times d}{(x-\mu)^T} \underset{d \times d}{\Sigma^{-1}} \underset{d \times 1}{(x-\mu)}\right)$$

$x \in \mathbb{R}^d$ x is $d \times 1$, μ $d \times 1$

$$f(x_1, \ldots, x_n; \mu, \Sigma) = \prod_{i=1}^{n} f(x_i; \mu, \Sigma)$$

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \hat{\mu} = \text{MLE of } \mu$$

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} \underset{d \times 1}{(x_i - \mu)} \underset{1 \times d}{(x_i - \mu)^T}$$

$$\left(= \frac{1}{n}\sum_{i=1}^{n} (x_i - \mu)^2 \quad \text{if } d=1 \right)$$

MIXTURE OF K NORMALS each d-dimensional

$$\theta = \begin{array}{lll} \mu_1 \in \mathbb{R}^d & \Sigma_1 \quad d \times d \text{ matrix} & \pi_1 \\ \mu_2 \in \mathbb{R}^d & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ \mu_K & \Sigma_K & \pi_K \end{array}$$

SEARCH FOR MAXIMA OF $L(\theta; x_1, \ldots, x_n)$

LAST TIME:

DERIVED A SYSTEM OF EQ'NS THAT A CRITICAL POINT SHOULD SATISFY

(1) $$\mu_j = \frac{1}{N_j} \sum_{i=1}^{n} \gamma^{(i)}(j)\, x_i \qquad j = 1, \ldots, K$$

$$N_j = \sum_{i=1}^{n} \gamma^{(i)}(j)$$

(2) $$\Sigma_j = \frac{1}{N_j} \sum_{i=1}^{n} \gamma^{(i)}(j) (x_i - \mu_k)(x_i - \mu_k)^T$$

(3) $$\pi_j = \frac{N_j}{n}$$

$$\gamma^{(i)}(j) = \mathbb{P}(Z_i = j \mid X_i; \mu_j, \Sigma_j)$$

Z_i is a hidden variable $Z_i \in \{1, 2, \ldots, k\}$

$X_i \mid Z_i = j$ is $N(\mu_j, \Sigma_j)$

$$\gamma^{(i)}(j) = \mathbb{P}\left(Z_i = j \mid X_i \,;\, \begin{array}{l} \mu_1, \ldots, \mu_K \\ \Sigma_1, \ldots, \Sigma_K \\ \pi_1, \ldots, \pi_K \end{array}\right)$$

$$= \frac{\mathbb{P}(X_i \mid Z_i = j)\,\mathbb{P}(Z_i = j)}{\sum_l \mathbb{P}(X_i \mid Z_i = l)\,\mathbb{P}(Z_i = l)}$$

$$= \frac{f(x_i; \mu_j, \Sigma_j)\,\pi_j}{\sum_l f(x_i; \mu_l, \Sigma_l)\,\pi_l}$$

$\theta^{(0)}$ initialize

$\theta^{(1)}$ updated estimate at θ; iterate

At each stage, $l(\theta^{(i)}) \geq l(\theta^{(i-1)})$