544NA
9/3/2013

TODAY: pdf ESTIMATION

#1: PARAMETRIC

① EXPONENTIAL FAMILIES

② PARAMETER ESTIMATION, ML,

~~ⓐ ML, ⓑ MAP~~

~~③ MIXTURE DENSITIES~~

MAKE SOME ASSUMPTIONS

DENSITY ESTIMATION

② NON-PARAMETRIC

ESTIMATION

DATA

"EMPIRICAL DISTRIBUTION"

$\boxed{p(x) \text{ IS "SMOOTH" W/SMOOTHNESS}}$ → $\boxed{\text{OPTIMAL SMOOTHING}}$ → $\hat{p}(x)$

$\mathcal{X} = \{x_1, \ldots, x_N\}$ → $\boxed{p_e(x) = \frac{1}{N}\sum_{i=1}^{N} \delta(x - x_i)}$

$x_n \in \mathcal{X}$

$\boxed{p(x) = e^{-g(x,\theta)}, \text{ FOR SOME KNOWN } g(x,\theta), \text{ UNKNOWN } \theta}$ → $\boxed{\text{ESTIMATE } \theta}$ → $\hat{p}(x)$

① PARAMETRIC

pdf: $\mathcal{X} \in \mathbb{R}^D$, UNCOUNTABLY INFINITE

pmf: $\mathcal{X}$ = CATEGORICAL OR INTEGER SET, FINITE OR COUNTABLY INFINITE

TODAY: PARAMETRIC

EXPONENTIAL FAMILIES: $p(x) = e^{-g(x,\theta)}$

PARAMETRIC ASSUMPTION: $g(x,\theta)$ KNOWN "DISCRIMINANT FUNCTION"

θ UNKNOWN "PARAMETERS"

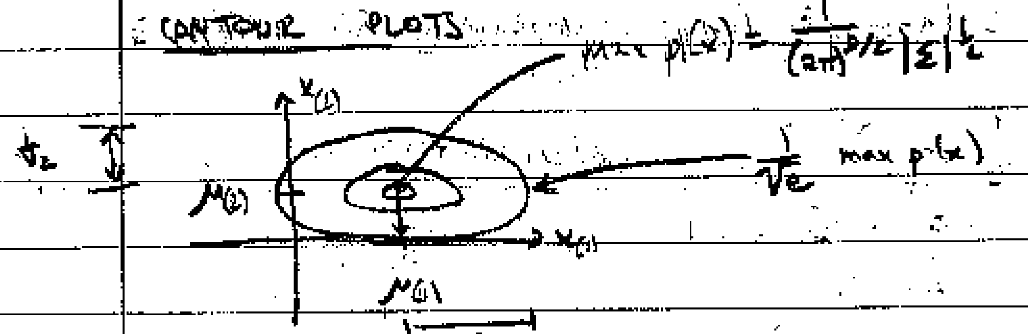
EXAMPLE 1: GAUSSIAN $p(x) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}$

$\boxed{|\Sigma| = \det(\Sigma)}$ $= e^{-g(x,\theta)}$

$$g(x,\theta) = \frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu) + \frac{1}{2}\ln|\Sigma| + \frac{D}{2}\ln(2\pi)$$

θ = PARAMETER VECTOR = $\begin{bmatrix} \mu \\ \Sigma(\cdot) \end{bmatrix}$ } D PARAMETERS; } ONLY $D(D+1)/2$ DISTINCT PARAMETERS BECAUSE Σ IS SYMMETRIC

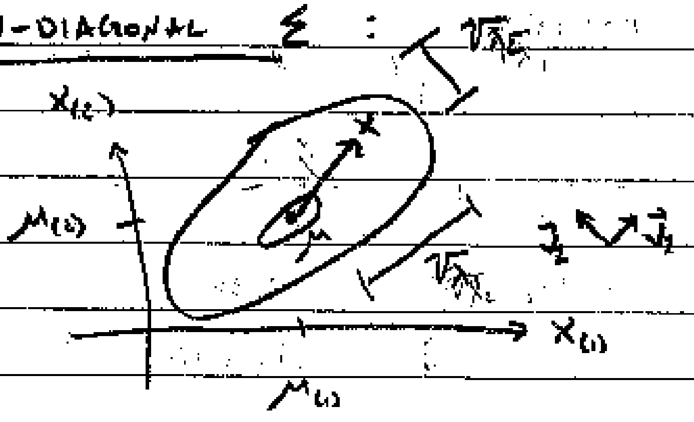
CONTOUR PLOTS



$\Sigma = \begin{bmatrix} \sigma_{(1)}^2 & 0 & 0 \\ 0 & \sigma_{(2)}^2 & 0 \\ 0 & 0 & \sigma_{(3)}^2 \end{bmatrix}$ DIAGONAL

$\Leftrightarrow$ PRINCIPAL AXES OF CONTOUR PLOT ARE PARALLEL TO $x_{(1)}$ $x_{(2)}$ AXES

NON-DIAGONAL Σ :



EIGENVECTORS $\vec{V}_d$, EIGENVALUES λ_d :

$$\Sigma \vec{v}_d = \vec{v}_d \lambda_d$$

$$\Sigma V = V \Lambda \qquad V = [\vec{v}_1, \dots, \vec{v}_D], \ \Lambda = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_D \end{bmatrix}$$

Σ POSITIVE DEFINITE $\Leftrightarrow$ $\lambda_d > 0$

EIGENVECTORS NORMALIZED SO $VV^T = V^TV = I$

$$\Rightarrow \boxed{\begin{aligned} &\Sigma = V\Lambda V^T \qquad \Lambda = V^T \Sigma V \\ &\Sigma^{-1} = V \Lambda^{-1} V^T = V \begin{bmatrix} \frac{1}{\lambda_1} & 0 & 0 \\ 0 & \frac{1}{\lambda_2} & 0 \\ 0 & 0 & \frac{1}{\lambda_D} \end{bmatrix} V^T \end{aligned}}$$

$$g(x) = \frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu) + \frac{1}{2} \ln |\Sigma| + \frac{D}{2} \ln(2\pi)$$

$$= \frac{1}{2}(x-\mu)^T V \Lambda^{-1} V^T (x-\mu) + \frac{1}{2} \ln |V \Lambda V^T| + \frac{D}{2} \ln(2\pi)$$

$$\boxed{g(x) = \frac{1}{2} \sum_{d=1}^{D} \frac{\| v_d^T (x-\mu) \|_2^2}{\lambda_d} + \frac{1}{2} \sum_{d=1}^{D} \ln \lambda_d + \frac{D}{2} \ln(2\pi)}$$

① FIND $\tilde{x} = x - \mu$

② PROJECT $y_d = v_d^T (x-\mu)$

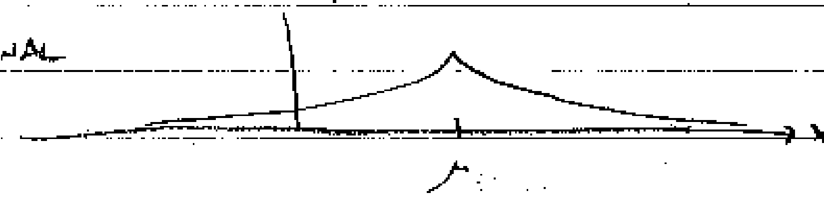
③ ADD

$$\boxed{g(x) = \frac{1}{2} \sum_{d=1}^{D} \left(\frac{\| y_d \|_2^2}{\lambda_d} + \ln \lambda_d \right) + \text{CONSTANT}}$$

EXAMPLE 2: LAPLACIAN

$$p(x) = \frac{1}{2\lambda} e^{-|x|/\lambda}$$

ONE-DIMENSIONAL

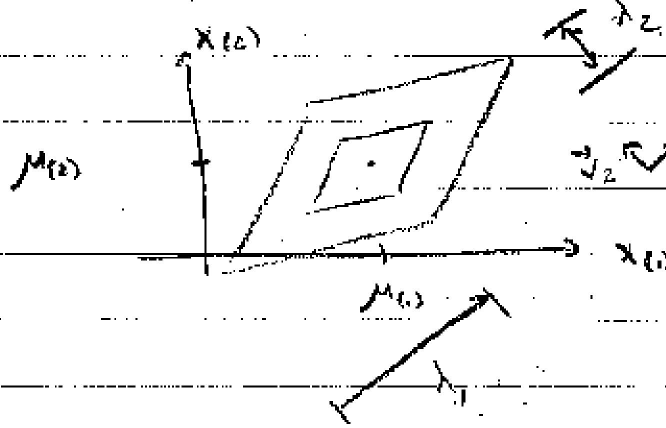


MULTI-DIMENSIONAL

$$V = [\vec{v}_1, \vec{v}_2, \dots \vec{v}_D] \qquad V^T V = V V^T = I$$

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_D \end{bmatrix} \qquad \text{ASSUME } \lambda_d > 0$$

THEN $p(x) = |2\Lambda|^{-1} e^{-\|\Lambda^{-1} V^T (x-\mu)\|_1}$



$$g(x) = \|\Lambda V^T (x-\mu)\|_1 + \ln|2\Lambda| = \sum_{d=1}^{D} \left(|v_d^T(x-\mu)| / \lambda_d + \ln 2\lambda_d \right)$$

$$\|x\|_1 \equiv \sum_{d=1}^{D} |x_{(d)}|$$

1. CENTER: $\tilde{x} = x - \mu$
2. PROJECT: $y_d = v_d^T (x-\mu)$
3. ADD: $g(x) = \sum_{d=1}^{D} \left(\frac{|v_d^T(x-\mu)|}{\lambda_d} + \ln \lambda_d \right)$ + CONSTANT

EXAMPLE 3: GENERALIZED MINKOWSKI pdf.

$$p(x) = \frac{1}{Z(\Delta)} e^{-\frac{1}{p}\|\Delta^{-\frac{1}{p}} V^T (x-\mu)\|_p^p}$$

$$g(x) = \frac{1}{p} \|\Delta^{-\frac{1}{p}} V^T (x-\mu)\|_p^p + \underbrace{\ln Z(\Delta)}_{\text{SOME NORMALIZER}}$$

$$= \frac{1}{p} \sum_{d=1}^{D} \frac{|v_d^T (x-\mu)|^p}{\lambda_d} + \ln Z(\Delta).$$

MINKOWSKI NORM: $\|y\|_p \equiv \sqrt[p]{\sum_{d=1}^{D} |y(d)|^p}$

$$\|y\|_p^p = \sum_{d=1}^{D} |y(d)|^p$$

EXAMPLE 4: MULTINOMIAL pmf

$$x_i \in \{1, \ldots, M\}$$

$$p(x_i) = \begin{cases} \pi_m & x_i = m \in \{1, \ldots, M\} \\ \emptyset & \text{ELSE} \end{cases}$$

$$\boxed{\sum_{m=1}^{M} \pi_m = 1, \quad \pi_m \geq 0}$$

EXPONENTIAL FAMILY: $p(x_i) = \prod_{m=1}^{M} \pi_m^{[x_i = m]}$

$$[x_i = m] = \begin{cases} 0 & x_i \neq m \\ 1 & x_i = m \end{cases}$$

SO $p(x_i) = e^{-\sum_{m=1}^{M} [x_i = m] \ln \frac{1}{\pi_m}} = e^{-z_i^T \theta}$ $\boxed{g(x) = \theta^T z(x)}$

$\theta = [\ln(\frac{1}{\pi_1}), \ldots, \ln(\frac{1}{\pi_M})]^T$ $\quad z_i = [0, 0, \ldots, \underset{m^{TH} \text{ DIMENSION}}{1}, \ldots, 0]$

MAXIMUM LIKELIHOOD

PARAMETER ESTIMATION: EXPONENTIAL FAMILY

$$\theta = \underset{\theta}{\operatorname{argmax}}\ \mathcal{L}(\mathbb{X}, \theta)$$

$$\mathcal{L}(\mathbb{X}, \theta) \equiv p(\mathbb{X} \mid \theta) = \prod_{n=1}^{N} p(x_n \mid \theta)$$

$$\theta = \operatorname{argmax}\ \ln \mathcal{L}(\mathbb{X}, \theta) = \operatorname{argmax} \sum_{n=1}^{N} \ln p(x_n \mid \theta)$$

$$\boxed{\theta = \operatorname{argmin} \sum_{n=1}^{N} g(x_n; \theta)}$$

SOLVE: $-\nabla_\theta \ln \mathcal{L}(\mathbb{X}, \theta) = 0$

$$\boxed{\sum_{n=1}^{N} \nabla_\theta g(x_n; \theta) = 0}$$

GAUSSIAN $\quad \sum_{n=1}^{N} \nabla_\mu g(x_n; \theta) = \Sigma^{-1} \left(\sum_{n=1}^{N} (\mu - x_n) \right)$

$$= 0 \quad \text{IFF} \quad \boxed{\hat{\mu} = \frac{1}{N} \sum_{n=1}^{N} x_n}$$

$$\sum_{n=1}^{N} \frac{\partial}{\partial \lambda_d} g(x_n, \theta) = \frac{1}{2} \sum_{n=1}^{N} \left(- \frac{(v_d^T(x_n - \mu))^2}{\lambda_d^2} + \frac{1}{\lambda_d} \right)$$

$$= 0 \quad \text{IFF} \quad \boxed{\hat{\lambda}_d = \frac{1}{N} \sum_{n=1}^{N} \left(v_d^T(x_n - \mu)\right)^2}$$

MORE GENERALLY $\quad \boxed{\hat{\Sigma} = \frac{1}{N} \sum_{n=1}^{N} (x_n - \mu)(x_n - \mu)^T}$

LAPLACIAN $\quad \boxed{\hat{\mu} = \operatorname{median}(\mathbb{X}), \quad \hat{\lambda}_d = \frac{1}{N} \sum_{n=1}^{N} \left| v_d^T(x_n - \mu) \right|}$

MULTINOMIAL $\quad \boxed{\hat{\pi}_m = \frac{1}{N} \sum_{n=1}^{N} [x_n = m] = \frac{\#\text{ OCCURRENCES } m}{\#\text{ OBSERVATIONS}}}$