ECE544NA
10/17/2013

TODAY: ERROR METRICS

1. MMSE ESTIMATION
   - REGRESSION
   - CLASSIFICATION
2. ML ESTIMATION
   - GAUSSIAN NOISE
   - L_p NOISE
3. pdf ESTIMATION
   - VARIANCE
   - GVGMM

MMSE ESTIMATION

$$y_k(x) = \underset{y(x) \in \mathcal{H}}{\operatorname{argmin}} \sum_{n=1}^{N} (y_k^n - t_k^n)^2$$

$$\mathcal{H} = \left\{ y_k(x) : y_k(x) = g\left(\sum_{j=0}^{M} w_{kj}\, g\left(\sum_{i=0}^{D} w_{ji} x_i\right)\right)\right\}$$

$$\lim_{\substack{N \to \infty \\ M \to \infty}} y_k(x) = \operatorname{argmin} \mathcal{E}_\infty$$

$$\mathcal{E}_\infty = \iint (y_k(x) - t_k)^2 p(x, t_k)\, dx\, dt$$

$$= \iint \Big\{ (y_k(x) - \langle t_k | x \rangle)^2 + (\langle t_k | x \rangle - t_k)^2 + 2 (y_k(x) - \langle t_k | x \rangle)\cancel{(t_k - \langle t_k | x \rangle)} \Big\} p(x, t_k)\, dx\, dt_k$$

$$E[t_k - \langle t_k | x \rangle \,|\, x] = 0$$

$$y_k(x) = \operatorname{argmin} \iint (y_k(x) - \langle t_k | x \rangle)^2 p(x, t_k)\, dx\, dt_k$$

$$\boxed{y_k(x) = \langle t_k | x \rangle}$$

AT THE MINIMUM,

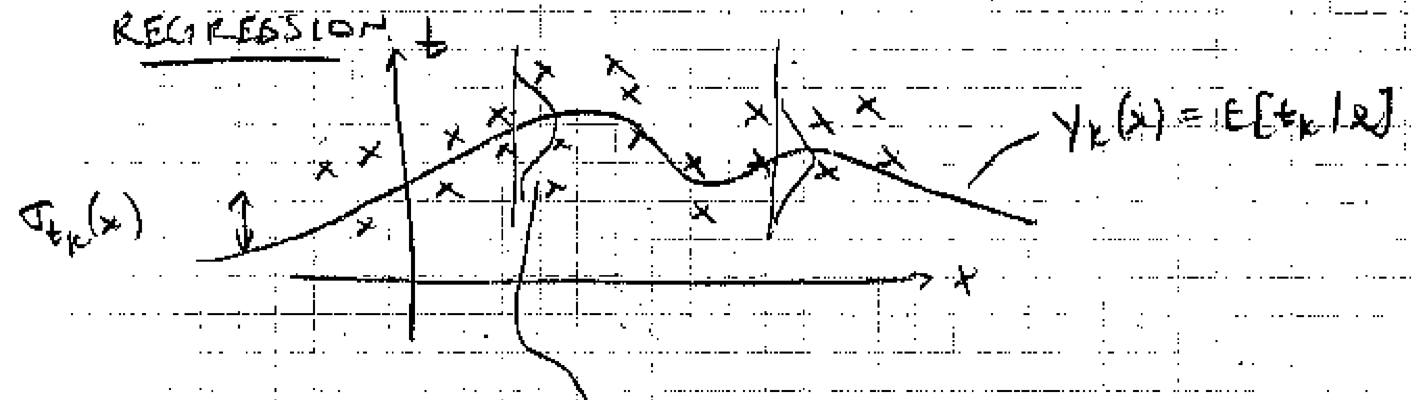
$$\mathcal{E}_\infty = \iint (t_k - \langle t_k | x \rangle)^2 p(x, t_k)\, dx\, dt_k$$

$$= \int \left[\int (t_k - \langle t_k | x \rangle)^2 p(t_k | x)\, dt_k \right] p(x)\, dx$$

$$\boxed{\mathcal{E}_\infty = \int \sigma_{t_k}^2(x)\, p(x)\, dx}$$

$$\boxed{\sigma_{t_k}^2(x) = E\left[(t_k - \langle t_k | x \rangle)^2 \,\middle|\, x \right]}$$

REGRESSION



$p(t_k | x)$ MODELED AS
$\mathcal{N}(t_k : \langle t_k | x \rangle, \sigma_{t_k}^2(x))$

CLASSIFICATION

$$t_k^n = \begin{cases} 1 & \text{IF } x^n \text{ IN CLASS } k \\ 0 & \text{IF } x^n \notin \text{ CLASS } k \end{cases}$$

$$t_k^n = \begin{cases} 1 & \text{w/ PROB } \Pr\{C_k | x^n\} \\ 0 & \text{w/ PROB } 1 - \Pr\{C_k | x^n\} \end{cases}$$

$$E[t_k | x] = \Pr\{C_k | x^n\}$$

$$\boxed{\lim_{\substack{N \to \infty \\ M \to \infty}} y_k(x) = \Pr\{C_k | x\}}$$

MAXIMUM LIKELIHOOD NEURAL NETS

ASSUME: $t_k^n = \underbrace{\langle t_k | x^n \rangle}_{\text{DETERMINISTIC TRUE EXPECTATION}} + \underbrace{v^n}_{\text{RANDOM ZERO-MEAN NOISE}}$

GAUSSIAN NOISE:

ASSUME $p_v(v^n) = \dfrac{1}{(2\pi\sigma^2)^{c/2}} e^{-\frac{1}{2\sigma^2}\|v^n\|^2}$

MODEL: $p(T|X,W) = \prod_{n=1}^{N} \prod_{k=1}^{c} \dfrac{1}{(2\pi\sigma^2)^{1/2}} e^{-\frac{1}{2\sigma^2}(t_k^n - y_k(x^n))^2}$

CHOOSE $W_{ML} = \arg\max p(T|X,W)$

$$= \arg\min \sum_{n=1}^{N} \sum_{k=1}^{c} -\ln p(t_k^n | x^n, W)$$

$$= \arg\min \sum_{n=1}^{N} \sum_{k=1}^{c} \frac{(t_k^n - y_k(x^n))^2}{2\sigma^2}$$

$$\boxed{W_{ML} = \arg\min \sum_{n=1}^{N} \sum_{k=1}^{c} (t_k^n - y_k(x^n))^2}$$

NORMALIZED MSE = A NORMALIZED METRIC OF NONLINEAR REGRESSION QUALITY

$$\mathcal{E}_{NMSE} = \frac{1}{c} \sum_{k=1}^{c} \frac{\sum_{n=1}^{N} (t_k^n - y_k(x^n))^2}{\sum_{n=1}^{N} (t_k^n - \bar{t}_k)^2}$$

$\mathcal{E}_{NMSE} = 0$ PERFECT

$\mathcal{E}_{NMSE} = 1$ NO BETTER THAN $y_k^n = \bar{t}_k$

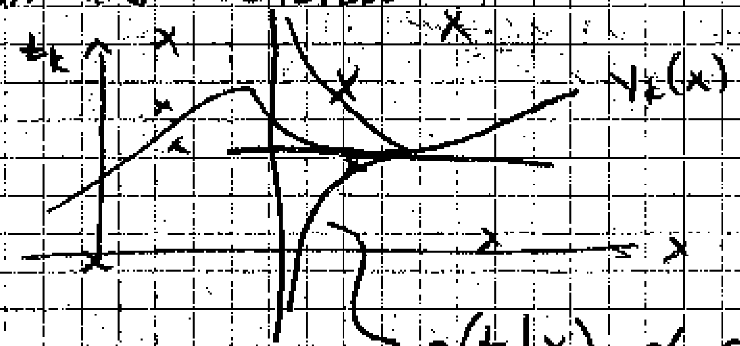
L_p NOISE

MODEL $\quad p(T|X,W) \propto \prod_{n=1}^{N} \prod_{k=1}^{c} e^{-\| t_k^n - y_k(x^n) \|^p}$

$$\mathcal{E} = -\ln p(T|X|W)$$

$$= \sum_{n=1}^{N} \sum_{k=1}^{c} |t_k^n - y_k(x^n)|^p$$

$p < 2$ IS BETTER THAN $p=2$ AT IGNORING OUTLIERS



$p(t|x) \propto e^{-|t_k - y_k(x)|}$

ERROR BACK-PROP

$$\frac{\partial \mathcal{E}}{\partial w_{ji}} = \sum_n \sum_k |y_k(x^n) - t_k^n|^{p-1} \operatorname{sign}(y_k(x^n) - t_k^n) \frac{\partial y_k^n}{\partial w_{ji}}$$

$$= \delta_j z_i$$

$$\delta_j = \sum_n \sum_k \delta_k w_{kj} g'(a_j)$$

$$\delta_k = \sum_n |y_k(x^n) - t_k^n|^{p-1} \operatorname{sign}(y_k(x^n) - t_k^n) g'(a_k)$$

pdf ESTIMATION

CAN WE USE MULTIPLE NNS TO ESTIMATE

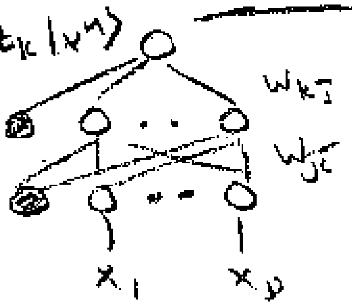
$p(t_k|x)$?

METHOD 1: CONDITIONAL VARIANCE

$\hat{y}_k^n \quad \tilde{y}_k^n$

MODEL: $\boxed{p(t_k|x) = \mathcal{N}(t_k : \langle t_k|x\rangle, \sigma^2_{t_k}(x))}$

$y_k^n \approx \langle t_k|x^n\rangle \longrightarrow (t_k^n - y_k^n)^2$ IS TARGET FOR



① TRAIN 1ST NETWORK s.t. $y_k \approx \langle t_k|x\rangle$

② TRAIN $\tilde{y}_k(x) = g\left(\sum_{j=0}^{M} \tilde{w}_{kj}\, g\left(\sum_{i=0}^{D} \tilde{w}_{ji} x_i\right)\right)$

TO MINIMIZE

$$\sum_{n=1}^{N}\left(\tilde{y}_k(x^n) - (t_k^n - y_k(x^n))^2\right)^2$$

$$\approx E\left[(\tilde{y}_k(x^n) - (t_k - \langle t_k|x^n\rangle)^2\right]$$

$$\Rightarrow \boxed{\tilde{y}_k^n \approx E\left[(t_k - \langle t_k|x\rangle)^2\right] = \sigma^2_{t_k}(x)}$$

METHOD 2: CONTINUOUSLY-VARIABLE-PARAMETER GAUSSIAN MIXTURE MODEL (CVGMM)

MODEL: $\boxed{p(t_k|x) = \sum_{j=1}^{M} \alpha_j(x) \frac{1}{\sqrt{2\pi\sigma_j^2(x)}} e^{-\frac{1}{2}\frac{(t_k^n - \mu_{jk}(x))^2}{\sigma_j^2(x)}}}$

NETWORK 1 NETWORK 2 NETWORK 3

$\longrightarrow$ HIDDEN NODES CALLED: z_j^{α} z_j^{σ} z_{jk}^{μ}

IMPOSING CONSTRAINTS BY NONLINEAR TRANSFORMATIONS

CONSTRAINT #1: $\sigma_j^2(x) > 0$

SOLUTION: $\sigma_j^2(x) = \exp(z_j^{\sigma})$

CONSTRAINT #2 : $0 \le \alpha_j(x) \le 1$

$$\sum_{j=1}^{m} \alpha_j(x_i) = 1$$

SOLUTION: SOFTMAX FUNCTION

$$\alpha_j(x) = \mathrm{softmax}(z_j^\alpha) = \frac{\exp(z_j^\alpha)}{\sum_{\ell=1}^{m} \exp(z_\ell^\alpha)}$$

$$\frac{\partial \alpha_\ell(x)}{\partial z_j^\alpha} = \begin{cases} \alpha_j - \alpha_j^2 & j = \ell \\ -\alpha_j \alpha_\ell & j \neq \ell \end{cases}$$

GRADIENT DESCENT

$$\mathcal{E} = -\sum_{n=1}^{N} \ln \left\{ \sum_{j=1}^{m} \alpha_j(x^n) \sum_{k=1}^{c} \mathcal{N}\left(t_k^n : \mu_{jk}(x^n), \sigma_j^2(x^n)\right) \right\}$$

ERROR BACK-PROP

$$\delta_j^\alpha = \frac{\partial \mathcal{E}}{\partial a_j^\alpha} = g'(a_j^\alpha) \frac{\partial \mathcal{E}}{\partial z_j^\alpha}$$

$$= g'(a_j^\alpha) \sum_{\ell=1}^{c} \frac{\partial \mathcal{E}}{\partial \alpha_\ell} \frac{\partial \alpha_\ell}{\partial z_j^\alpha}$$

$$= g'(a_j^\alpha) \sum_{\ell=1}^{c} \left(-\frac{\pi_\ell}{\alpha_\ell}\right) \left(\delta(j=\ell)\alpha_\ell - \alpha_j \alpha_\ell\right)$$

$$\boxed{\delta_j^\alpha = g'(a_j^\alpha)(\alpha_j - \pi_j)}$$

$$\pi_j = \frac{\alpha_j \mathcal{N}(t \mid \mu_j, \sigma_j^2)}{\sum_{\ell=1}^{m} \alpha_\ell \mathcal{N}(t \mid \mu_\ell, \sigma_\ell^2)}$$

$$\frac{\partial \mathcal{E}}{\partial \sigma_j} = -\pi_j \left(\frac{\|\vec{t} - \vec{\mu}_j\|^2}{\sigma_j^3} - \frac{c}{\sigma_j} \right)$$

$$\frac{\partial \sigma_j}{\partial z_j^\sigma} = \sigma_j$$

$$\boxed{\frac{\partial \mathcal{E}}{\partial z_j^\sigma} = -\pi_j \left(\frac{\|\vec{t} - \vec{\mu}_j\|^2}{\sigma_j^2} - c \right)}$$

$$\boxed{\frac{\partial \mathcal{E}}{\partial z_{jk}^\mu} = \pi_j \left(\frac{\mu_{jk} - t_k}{\sigma_j^2} \right)}$$