ECE544NA

10/22/2013

TODAY: INFORMATION THEORY

- BLAHUT ERROR EXPONENT
- MINIMUM CROSS ENTROPY TRAINING

A BINARY HYPOTHESIS TESTER ANN

GIVEN: $k^n \in \{1, \ldots, D\}$ ONE-OF-D

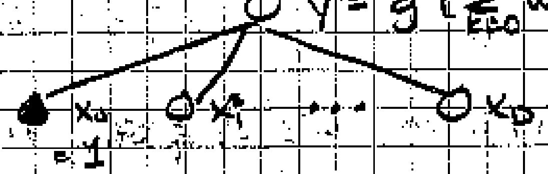
TARGET: $t^n = 1$ OR $t = 2$

KNOWN: $q_{1k} = \Pr\{k^n = k \mid t^n = 1\}$, $q_{2k} = \Pr\{k^n = k \mid t^n = 2\}$

NN ENCODING
$$x_d^n = \begin{cases} 1 & k^n = d \\ 0 & \text{ELSE} \end{cases}$$

i.e. $\vec{x}^n \in \{ \{0,1\}^D : \|\vec{x}\| = 1 \}$

$$Y = g\left(\sum_{k=0}^{D} w_k x_k\right) \qquad g(a) = \begin{cases} 2 & a > 0 \\ 1 & a < 0 \end{cases}$$



ML DECISION RULE: CHOOSE

$y = 2$ IFF k: $\ln \frac{q_{2k}}{q_{1k}} > 0$

ASSUME $q_{1k} > 0$, $q_{2k} > 0$ FOR ALL k

$$\Rightarrow \quad w_k = \begin{cases} \ln \frac{q_{2k}}{q_{1k}} & 1 \le k \le D \\ 0 & k = 0 \end{cases}$$

MAP DECISION RULE: CHOOSE $y = 2$ IFF $\ln \frac{\Pr\{C_2\} q_{2k}}{\Pr\{C_1\} q_{1k}} > 0$

$$\Rightarrow \quad w_k = \begin{cases} \ln \frac{q_{2k}}{q_{1k}} & 1 \le k \le D \\ \ln \frac{\Pr\{C_2\}}{\Pr\{C_1\}} & k = 0 \end{cases}$$

$\ln \frac{q_{2k}}{q_{1k}} > \ln \frac{\Pr(C_1)}{\Pr(C_2)}$

GENERAL NEYMAN-PEARSON TEST

CHOOSE $y=2$ IFF $\ln \frac{q_{2k}}{q_{1k}} > \Theta$

$$W_k = \begin{cases} \ln \frac{q_{2k}}{q_{1k}} & 1 \le k \le D \\ \Theta & k=0 \end{cases}$$

TYPE 1 ERROR: $t=2$ BUT $y=1$

$$\alpha = \Pr\{y=1 \mid t=2\}$$

$$= \sum_k q_{2k}\, u\left(\ln \frac{q_{2k}}{q_{1k}} - \Theta\right)$$

TYPE 2 ERROR $\beta = \Pr\{y=2 \mid t=1\}$

$$= \sum_k q_{1k}\, u\left(\Theta - \ln \frac{q_{2k}}{q_{1k}}\right)$$

KULLBACK-LEIBLER DISTORTION

$$J(q_1; q_2) = \sum_k q_{1k} \ln \frac{q_{1k}}{q_{2k}} = \sum_k q_{1k} \ln q_{1k} - \sum_k q_{1k} \ln q_{2k}$$

- CONVEX IN BOTH ITS ARGUMENTS:

$$\frac{\partial^2 J}{\partial q_{2\ell}^2} = \frac{q_{1\ell}}{q_{2\ell}^2} > 0$$

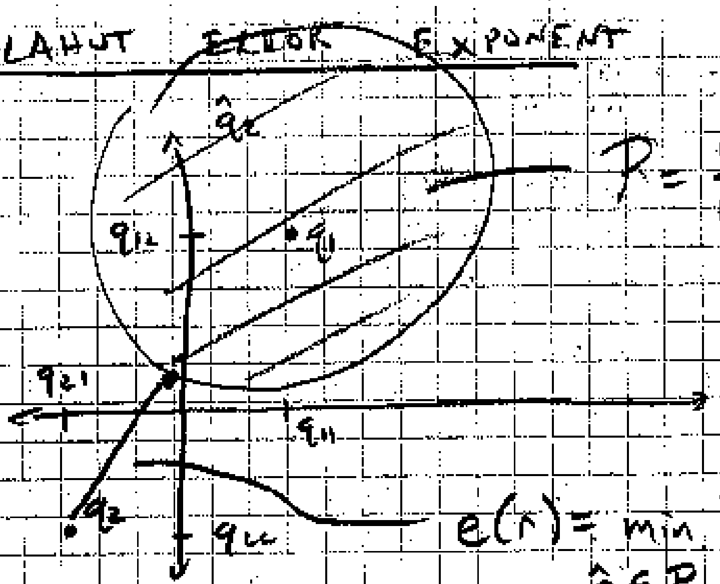
$$\frac{\partial^2 J}{\partial q_{1\ell}^2} = \frac{1}{q_{1\ell}} \ln \frac{1}{q_{2\ell}} > 0$$

- GLOBAL MINIMUM AT $q_{1k} = q_{2k}$ IS

$$J(q_1; q_2) = 0$$

- $J(q_1; q_2) > 0$ FOR all $q_1 \neq q_2$

BLAHUT ERROR EXPONENT



$$P = \{\hat{q} : J(\hat{q}; q_1) \le r\}$$

$$e(r) = \min_{\hat{q} \in P} J(\hat{q}; q_2)$$

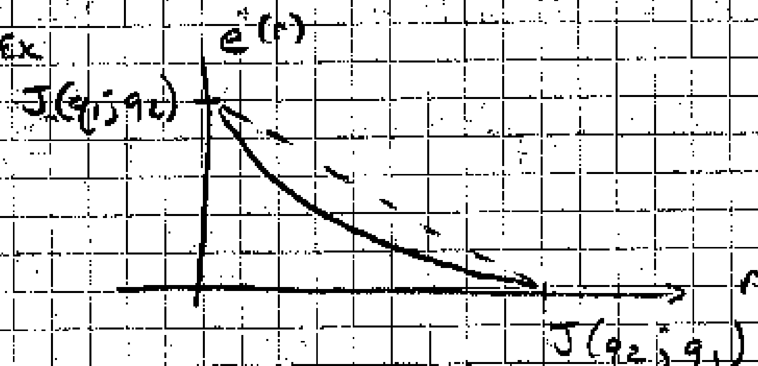
① e(r) IS EASY TO CALCULATE:

$$q^* = \arg\min_{\hat{q} \in P} J(\hat{q}; q_2)$$

IS GIVEN BY $$q^*_k = \frac{(q_{2k}\, q_{1k}^{s})^{\frac{1}{1+s}}}{\sum_{l} (q_{2k}\, q_{1k}^{s})^{\frac{1}{1+s}}}$$

FOR SOME $0 < s < \infty$

② e(r) IS CONVEX AND NON-NEGATIVE:



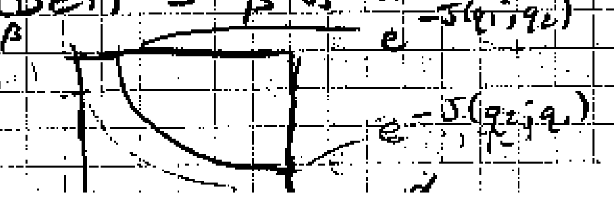
③ e(r) DETERMINES α, β, AND THE NN BIAS

CHOOSE $y = 2$ IFF $\ln \frac{q_{2k}}{q_{1k}} \ge r - e(r)$

$$\Longrightarrow \boxed{\begin{array}{l} \beta \le e^{-r} \\ \alpha \le e^{-e(r)} \end{array}}$$

DETECTION ERROR TRADEOFF (DET) $\equiv \beta$ VS α

$$W_k = \begin{cases} \ln \frac{q_{2k}}{q_{1k}} & 1 \le k \le D \\ r - e(r) & k = 0 \end{cases}$$



EER: $\alpha = \beta$

MAP: $r = \arg\min (\alpha P(C_1) + \beta P(C_2))$

K-L DIVERGENCE AS A TRAINING CRITERION

SUPPOSE $\vec{x}^n$ ARBITRARY, BUT

$$t_k^n = \begin{cases} 1 & \vec{x}^n \in C_k \\ 0 & \text{ELSE} \end{cases}$$

GOAL:

$$y_k^n = P_r\{C_k | \vec{x}^n\}$$

I.E. $p(\vec{t}^n | \vec{x}^n) = \prod_{k=1}^{c} (y_k^n)^{t_k^n}$

$$\boxed{\mathcal{E} = -\sum_{n=1}^{N} \ln p(\vec{t}^n | \vec{x}^n) = -\sum_{n=1}^{N} \sum_{k=1}^{c} t_k^n \ln y_k^n}$$

REMINDER: $\sum_k t_k^n = 1$

CHOOSE $\sum_k y_k^n = 1$, E.G. $y_k^n = \dfrac{\exp(a_k^n)}{\sum_{k=1}^{c} \exp(a_k^n)}$

THEN

$$J(t^n; y^n) = \sum_k t_k^n \ln\left(\frac{t_k^n}{y_k^n}\right)$$

$$= H(t^n; y^n) - H(t^n)$$

ENTROPY $H(t^n) \equiv -\sum_{k=1}^{c} t_k^n \ln t_k^n$

MINIMUM: $t_k^{n*} = \delta[k - k^*] = \begin{cases} 1 & k = k^* \\ 0 & \text{else} \end{cases}$

$H(t^n) = 0$

MAXIMUM: $t_k^n = \frac{1}{c}$ UNIFORM pmf

$H(t^n) = \ln\left(\frac{1}{c}\right)$

CROSS-ENTROPY $H(t^n; y^n) \equiv -\sum_{k=1}^{c} t_k^n \ln y_k^n$

MINIMUM: $y_k^n = t_k^n$, $H(t; y) = H(t)$

MAXIMUM: $y_k^n = \delta[k - k^*] \neq t_k^n$

$\Rightarrow H(t; y) = \infty$

LET'S CHOOSE $y_k^n = \underset{\ell}{\text{softmax}}\left(\sum_{j=0}^{M} w_{\ell j}\, g\left(\sum_{i=0}^{D} w_{ji}\, x_i^n\right)\right)$

S.T.

$$W = \operatorname{argmin} \sum_{n=1}^{N} J(\vec{t}_n, \vec{y}_n)$$

INDEP OF W

$$= \operatorname{argmin} \sum_{n=1}^{N} \left[H(t_n; y_n) - H(t_n)\right]$$

$$= \operatorname{argmin} \sum_{n=1}^{N} H(t_n; y_n)$$

$$= \operatorname{argmin}\left(-\sum_{n=1}^{N}\sum_{k=1}^{c} t_k^n \ln y_k^n\right)$$

<u>BACK-PROP</u>

$$\frac{\partial \mathcal{E}}{\partial w_{kj}} = \delta_k z_j$$

$$\delta_k \equiv \frac{\partial \mathcal{E}}{\partial a_k} = \sum_{n=1}^{N} \frac{\partial \mathcal{E}^n}{\partial a_k^n}$$

$$\mathcal{E}^n = -\sum_{k=1}^{c} t_k^n \ln y_k^n = -\sum_{k=1}^{c} t_k^n \ln\left(\frac{\exp(a_k^n)}{\sum_{\ell=1}^{c} \exp(a_\ell^n)}\right)$$

$$\frac{\partial \mathcal{E}^n}{\partial a_k^n} = \sum_{\ell=1}^{c} \frac{\partial \mathcal{E}^n}{\partial y_\ell^n}\frac{\partial y_\ell^n}{\partial a_k^n}$$

$$\frac{\partial \mathcal{E}^n}{\partial y_\ell^n} = -\frac{t_\ell^n}{y_\ell^n}$$

$$\frac{\partial y_\ell^n}{\partial a_k^n} = \begin{cases} y_k^n - (y_k^n)^2 & k = \ell \\ -y_k^n y_\ell^n & k \neq \ell \end{cases}$$

$$\boxed{\frac{\partial \mathcal{E}^n}{\partial a_k^n} = y_k^n - t_k^n}$$

$$\boxed{\delta_k = \sum_{n=1}^{N}(y_k^n - t_k^n)}$$

SAME AS $\frac{\partial \mathcal{E}}{\partial a_k}$ FOR $\mathcal{E} = \frac{1}{2}(t - y)^2$, $y(a) = a$

ONE MORE ABOUT CROSS-ENTROPY - IT'S ADDITIVE

LET $T = \begin{bmatrix} \vec{t}^{1T} \\ \vdots \\ \vec{t}^{NT} \end{bmatrix}$ $\quad Y = \begin{bmatrix} \vec{y}^{1T} \\ \vdots \\ \vec{y}^{NT} \end{bmatrix}$

LET $p(C_{k_n} | x^n) = t^n_{k_n}$ $\quad = p($

$\hat{p}(C_{k_n} | x^n) = y^n_{k_n}$

THEN $p(C_{k_1}, \ldots, C_{k_N} | x^1, \ldots, x^N) = t^1_{k_1} \cdots t^N_{k_N}$

$$H(T;Y) = -\sum_{k_1} \cdots \sum_{k_N} p(C_{k_1}, \ldots, C_{k_N} | x^1, \ldots, x^N) \ln \hat{p}(C_{k_1} \ldots$$

$$= -\sum_{k_1} \cdots \sum_{k_N} \left(\prod_{n=1}^{N} t^n_{k_n} \right) \ln \left(\prod_{n=1}^{N} y^n_{k_n} \right)$$

$$= -\sum_{k_1} \cdots \sum_{k_N} \left(\prod_{m=1}^{N} t^m_{k_m} \right) \left(\sum_{n=1}^{N} \ln y^n_{k_n} \right)$$

$$= -\sum_{k_1} t^1_{k_1} \ln y^1_{k_1} \sum_{k_2} \cdots \sum_{k_N} \left(\prod_{m=2}^{N} t^m_{k_m} \right)$$

$$- \sum_{k_1} t^1_{k_1} \sum_{k_2} t^2_{k_2} \ln y^2_{k_2} \sum_{k_3} \cdots \sum_{k_N} \left(\prod_{m=3}^{N} t^m_{k_m} \right)$$

$$= -\sum_{n=1}^{N} \sum_{k=1}^{C} t^n_k \ln y^n_k$$