PROBABILISTIC DISCRIMINATIVE MODELS (CH. 4.3)

DIRECT OPTIMIZATION OF $p(C_k|x)$ AS GENERALIZED LINEAR MODEL

$$p(C_k|\phi(x)) = y(\phi(x)) = \sigma(\underline{w}^T\underline{\phi})$$

BASIS FUNCTION < NONLINEAR FIXED / DATA DEPENDENT $\quad \phi(\underline{x})$

SEPARATION OF CLASSES LINEAR < IN INPUT SPACE / IN FEATURE SPACE $\updownarrow$? $\quad$ (FIG. 4.12)

LOGISTIC (REGRESSION) (CH. 4.3.2)

$$p(C_1|\underline{\phi}) = y(\underline{\phi}) = \sigma(\underline{w}^T\underline{\phi})$$

M-DIM FEATURE SPACE
M PARAMETERS

$$p(C_2|\underline{\phi}) = 1 - p(C_1|\underline{\phi})$$

LOGISTIC SIGMOID

(FOR ML FITTING
2M + M(M+1)/2 PARAMETERS!)

$\{\phi_n, t_n\}$, $t_n \in 0,1$ $\quad \phi_n = \phi(x_n) \quad n = 1 \ldots N \quad \underline{t} = (t_1, \ldots t_N)^T$

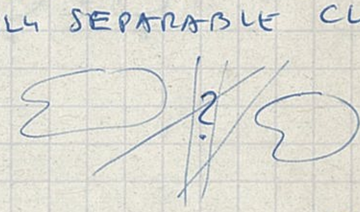
$$y_n = p(C_1|\phi_n)$$

$$p(\underline{t}|\underline{w}) = \prod_{n=1}^{N} y_n^{t_n}(1-y_n)^{1-t_n} \qquad y_n = \sigma(a_n) \qquad a_n = \underline{w}^T\phi_n$$

$$E(\underline{w}) = -\ln p(\underline{t}|\underline{w}) = -\sum_{n=1}^{N}\{t_n \ln y_n + (1-t_n)\ln(1-y_n)\}$$

$$\nabla E(\underline{w}) = \sum_{n=1}^{N}(y_n - t_n)\phi_n \qquad \leftarrow \qquad \frac{d\sigma}{da} = \sigma(1-\sigma)$$

→ SEQUENTIAL ALGORITHM
INITIAL CONDITIONS } ML OVERFITTING FOR LINEARLY SEPARABLE CLASSES



→ NO CLOSED FORM SOLUTION (σ!)
NON QUADRATIC
———
CONCAVE, UNIQUE MINIMUM

MAP!
(REGULARIZATION)

ITERATIVE REWEIGHTED LEAST SQUARES (NEWTON-RAPHSON) (CH. 4.3.3)

$$\underline{w}^{(new)} = \underline{w}^{(old)} - \underline{\underline{H}}^{-1}\nabla E(\underline{w}) \qquad \underline{\underline{H}} \text{ HESSIAN}, \frac{\partial^2}{\partial\partial}$$