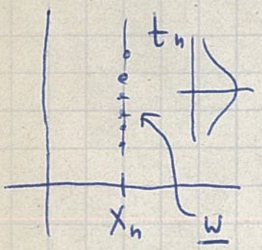
MAXIMUM LIKELIHOOD PRINCIPLE (BISHOP 3.1.1)



+ RANDOM ERROR

GIVEN: $\underline{x} = (x_1, \dots, x_N)^T$

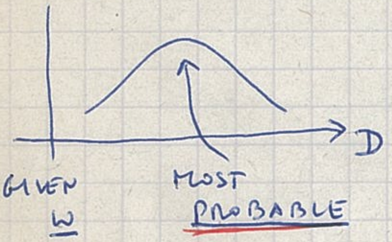
$\underline{t} = (t_1, \dots, t_N)^T$

ERROR MODELING AS: $p(t_n | \underline{w})$ E.G. $N(0, \sigma_n^2)$

ALL DATA: $p(D|\underline{w}) = p(t_1, \dots, t_N | \underline{w}) = L(\underline{w})$

AS FUNCTION OF D

= PROBABILITY OF OBSERVED DATA



AS FUNCTION OF $\underline{w}$

= FUNCTION (LIKELIHOOD) OF LIKELY PARAMETERS GENERATING THE DATA

$$\underline{w}^*_{ML} = \arg\max_{\underline{w}} L(\underline{w}) = \arg\min_{\underline{w}} \Big[\underbrace{-\ln L(\underline{w})}_{\ell(\underline{w})} \Big]$$

INDEPENDENT DATA + GAUSSIAN ERROR

$$p(t_n|\underline{w}) \sim N(y(x_n, \underline{w}), \sigma_e^2)$$ SAME NOISE FOR EVERY DATA

$$= \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma_e} \exp\left[-\frac{1}{2\sigma_e^2} \left(y(x_n, \underline{w}) - t_n \right)^2 \right]$$

$$p(D|\underline{w}) = \underbrace{\prod_{n=1}^{N} p(t_n|\underline{w})}$$

$$\ell(\underline{w}) = \frac{N}{2} \ln(2\pi) + \frac{N}{2} \ln \sigma_e^2 + \frac{1}{2\sigma_e^2} \sum_{n=1}^{N} \left(y(x_n, \underline{w}) - t_n \right)^2$$