$$\boxed{\underline{w}^*_{ML} = \left(\underline{\underline{\Phi}}^T \underline{\underline{\Phi}}\right)^{-1} \underline{\underline{\Phi}}^T \underline{t}}$$ $\left(= \underline{w}^*_{LS} \text{ IF INDEPENDENT + GAUSSIAN} \right)$

– SAME PROBLEM WITH OVERFITTING → REGULARIZATION !

– DATA NOT INDEPENDENT :

$$p(D \mid \underline{w}) = \frac{1}{(2\pi)^{N/2} \left|\Sigma_{ee}\right|^{1/2}} \exp\left[-\frac{1}{2} \left(\underline{t} - \underline{\underline{\Phi}}\,\underline{w}\right)^T \underline{\underline{\Sigma}}_{ee}^{-1} \left(\underline{t} - \underline{\underline{\Phi}}\,\underline{w}\right) \right]$$

$$\downarrow$$

$$\boxed{\underline{w}^*_{ML} = \left(\underline{\underline{\Phi}}^T \underline{\underline{\Sigma}}_{ee}^{-1} \underline{\underline{\Phi}}\right)^{-1} \underline{\underline{\Phi}}^T \underline{\underline{\Sigma}}_{ee}^{-1} \underline{t}}$$ (MARKOV - GAUSSIAN ESTIMATOR)

BIAS / VARIANCE TRADE-OFF (BISHOP 3.2)

MSE : NOISE VARIANCE + BIAS^2 + VARIANCE

ACHIEVABLE MIN

NOISE VARIANCE: $\mathcal{E}\{ (t_n(x_n) - t(x_n))^2 \}$ (TRUE)

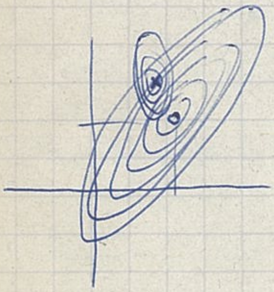
BIAS: $t(x_n) - \mathcal{E}\{ y(x_n, \underline{w}) \}$

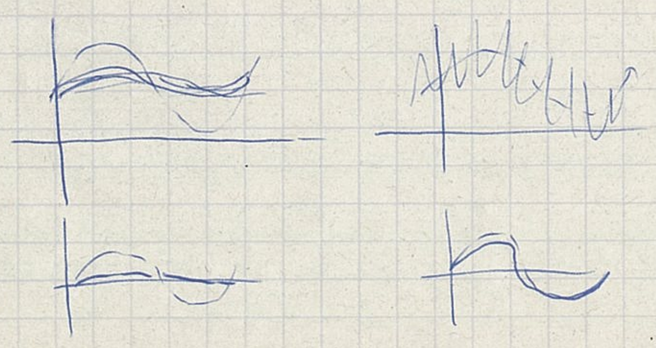
VARIANCE: $\mathcal{E}\{ (y(x_n, \underline{w}) - \mathcal{E}\{ y(x_n, \underline{w}) \})^2 \}$

(BISHOP CH1, 3)

$\lambda = \phi$ NO REGULARIZATION – NO BIAS LARGE VARIANCE

$\lambda \gg 1$ STRONG REGULARIZATION – LARGE BIAS SMALL VARIANCE





(BISHOP FIG 3.5)