

$$J(w) = \frac{1}{2N} \sum_{n=1}^N \left(\underbrace{w^T x^{(n)} - t_n}_{w_0 + w_1 x_n + \dots + w_M x_n^M} \right)^2 + \frac{\lambda}{2} \|w\|^2$$

$$[w_0, w_1, \dots, w_M]^T$$

$$x^{(n)} = [1, x_n, x_n^2, \dots, x_n^M]$$

$$\sum_{j=0}^M w_j^2$$

hyper-parameter $\lambda \geq 0$

$$\hat{w} = \arg \min_w J(w) \quad \frac{\partial J}{\partial w} = 0$$

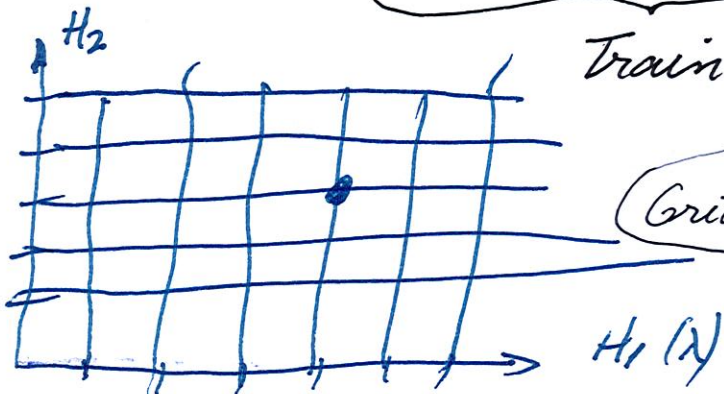
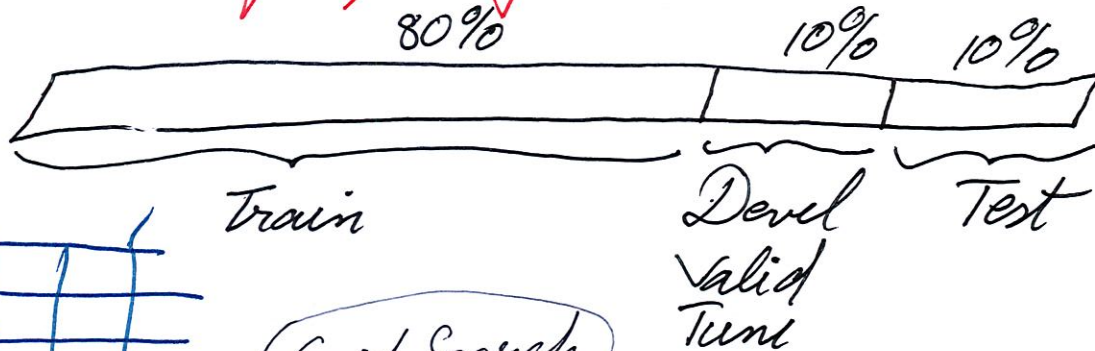
$$\ln \lambda = -18 \Rightarrow \lambda = e^{-18} = \frac{1}{e^{18}}$$

$$e = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n$$

$\hookrightarrow 2.718\dots$ Euler's number. $e^{-1} = 1/e$

Why not $\hat{w}, \hat{\lambda} = \arg \min_{w, \lambda} J(w, \lambda) \Rightarrow \hat{\lambda} = 0!$

Dataset D



Grid Search