

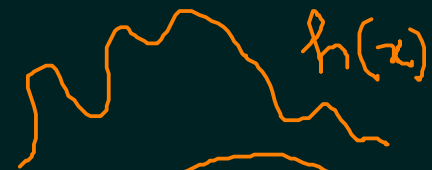
Core Learning Principles:

[OCCUM
RAZOR]

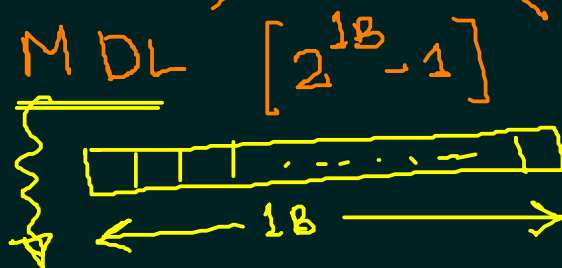
↳ Simple Hypothesis \rightsquigarrow Generalization

What is simple?

Simple is better why?



↳ Complexity of "h": order of poly



↳ Complexity of "H": (mathematical)

" $2^{1B} - 1$ "

↳ VC-Dimension, Entropy



↳ What is link??

$h \Rightarrow$ l bits \rightsquigarrow one among 2^l elements of H

Real-valued Parameters } $h \in$ 17th order Poly

$$a_0 + a_1x + \dots + a_{16}x^{16} + a_{17}x^{17}$$

"one among many"

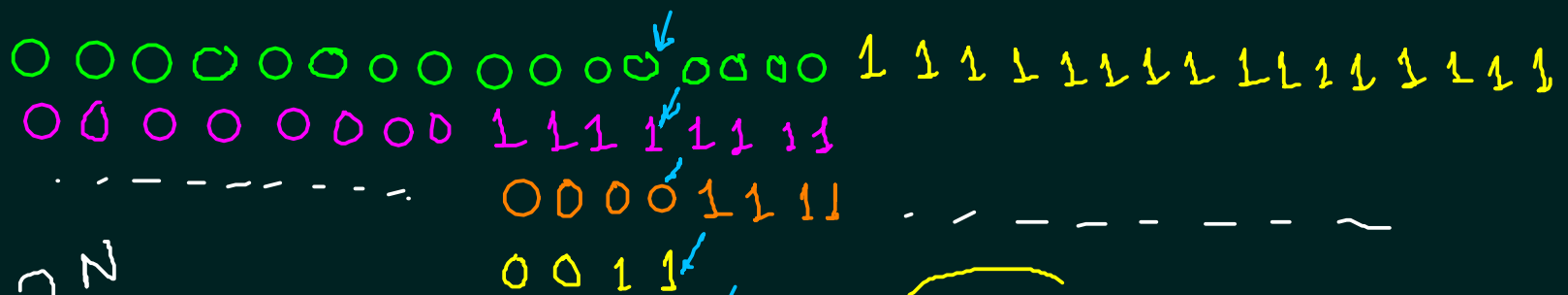
Exception:

$$H = \{ \}$$

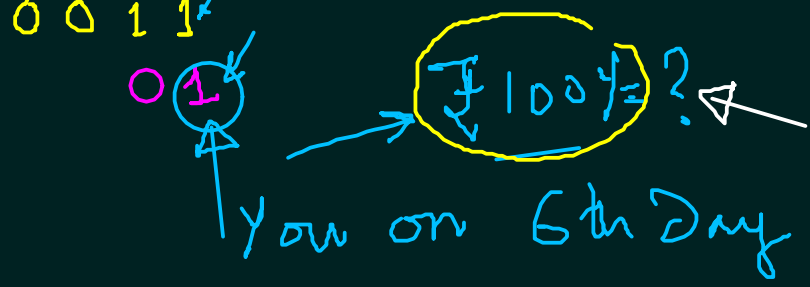
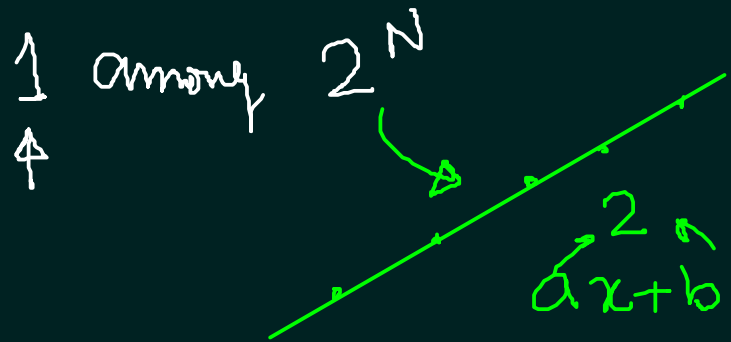
$$h = \langle sv \rangle$$



5 days ①



○
1
○
1
1

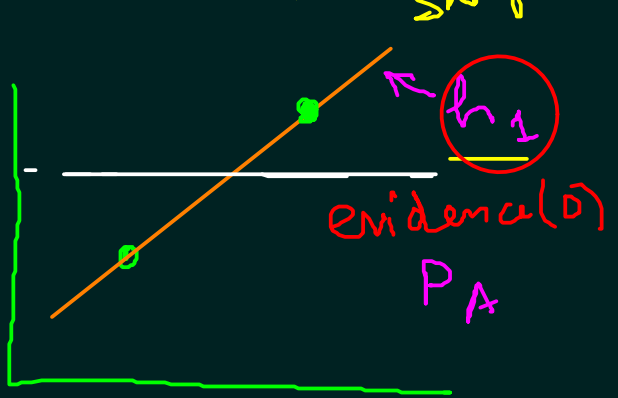


→ Fewer Simple Hyp than Complex ones: $m_{H_{MC}}(N) / 2^N$

(h) You ← $1 / 2^N$

Simple is better when it performs well also in out-of-sample

Rep. of out of sample fitting



"falsifiability"
↳ Evidence
← ?

Sampling Bias

1948 ~ Truman vs. Dewey

$[100]$ \Rightarrow

$$P[|E_{in} - E_{out}| > \epsilon] \leq \delta$$

↑ Wins (Prediction)



If data is sampled biased by Learning will o/p Biased "h"

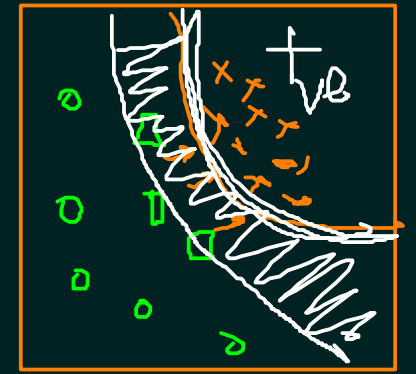
$w_1 D_1, w_2 D_2$

age \rightarrow
sal \rightarrow
years \rightarrow

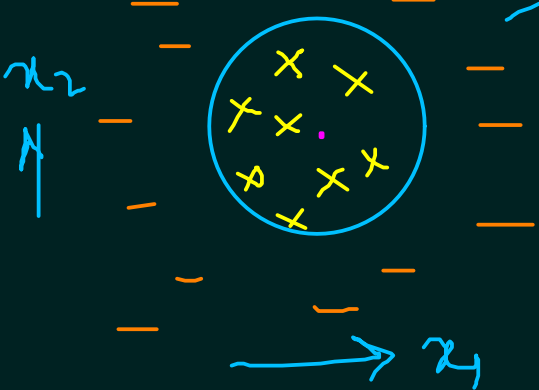
Credit
App/Deny

M/L
S/W

Accept



Data Snooping



$$Z = (1, x_1, x_2, x_1 x_2, x_1^2, x_2^2)$$

$$Z = (1, x_1^2, x_2^2)$$

$$Z = (1, x_1^2 + x_2^2)$$

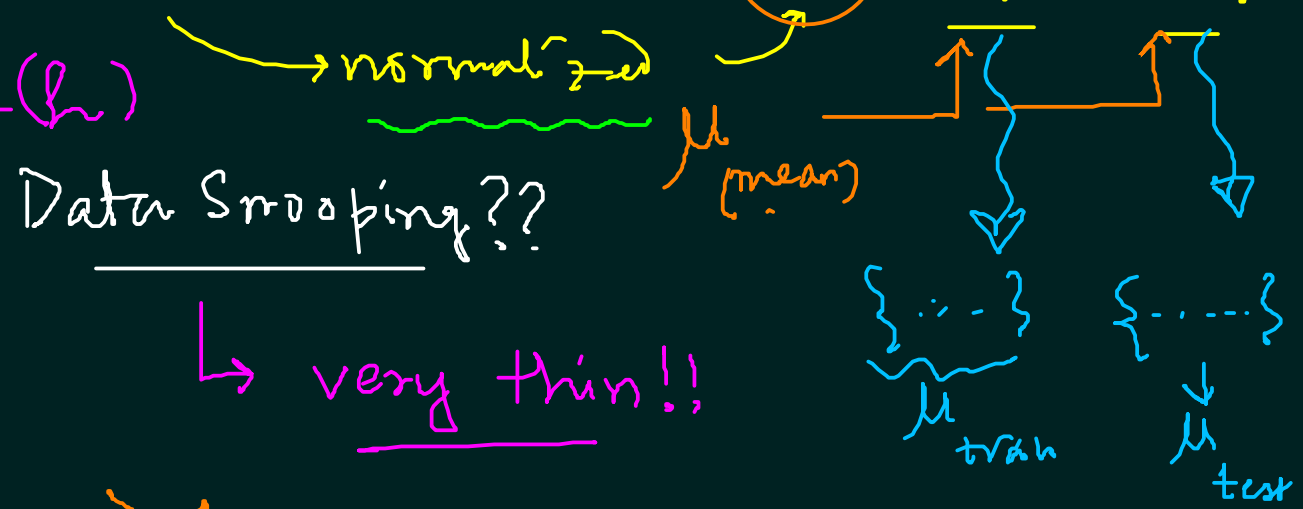
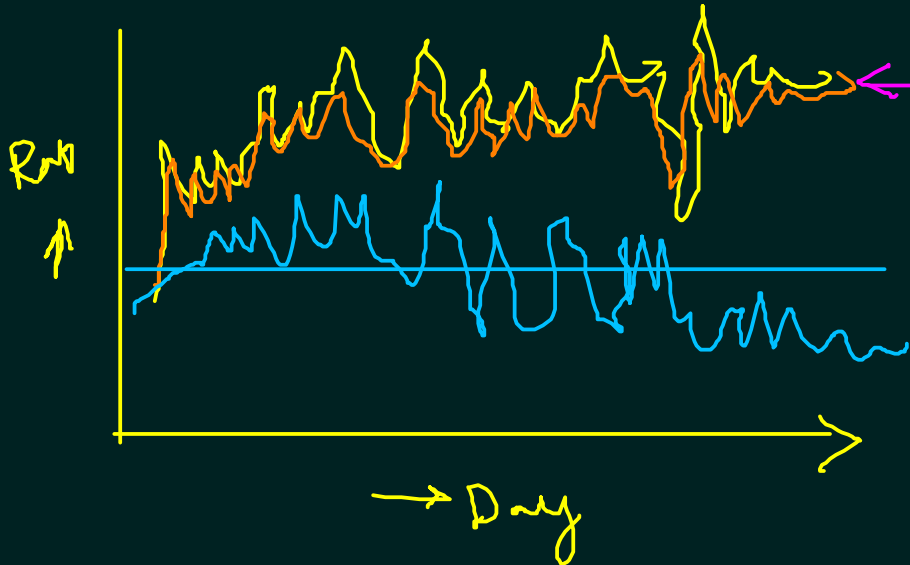
$$Z = (x_1^2 + x_2^2 - 0.6)$$

1 param

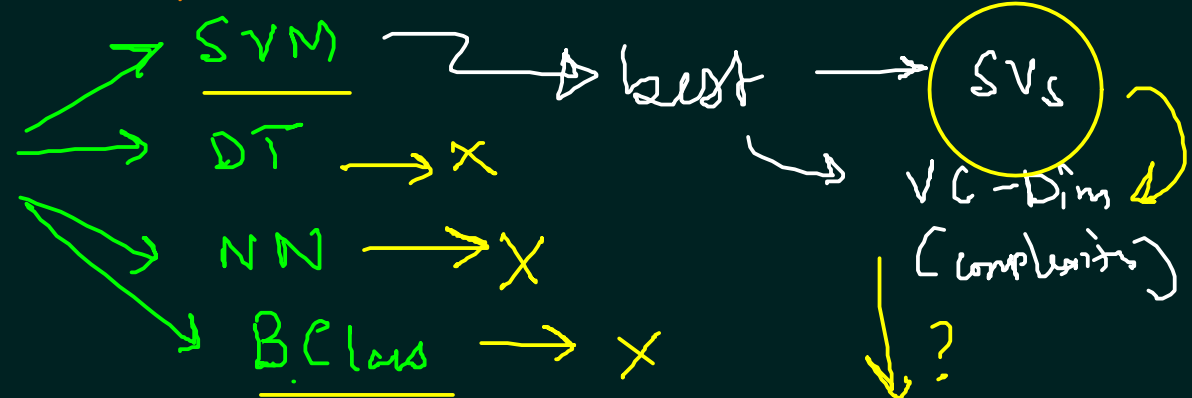
A

$\gamma_{\Delta-20}, \gamma_{\Delta-19}, \gamma_{\Delta-18} \dots \gamma_{\Delta-1} \rightarrow \Δ

$\mathcal{D} \neq \mathcal{D}_{Tr} \cup \mathcal{D}_{Test}$



Trying different Model



Ω

(2)

VC-dim $\mathcal{D} = \{SVM, DT, NN, BClass\}$

2 Remedies: \rightarrow Discipline, cross-check

\rightarrow Account for how much?

50 yr.

\rightarrow Age \uparrow 500

??