Design and Analysis of ML Experiments

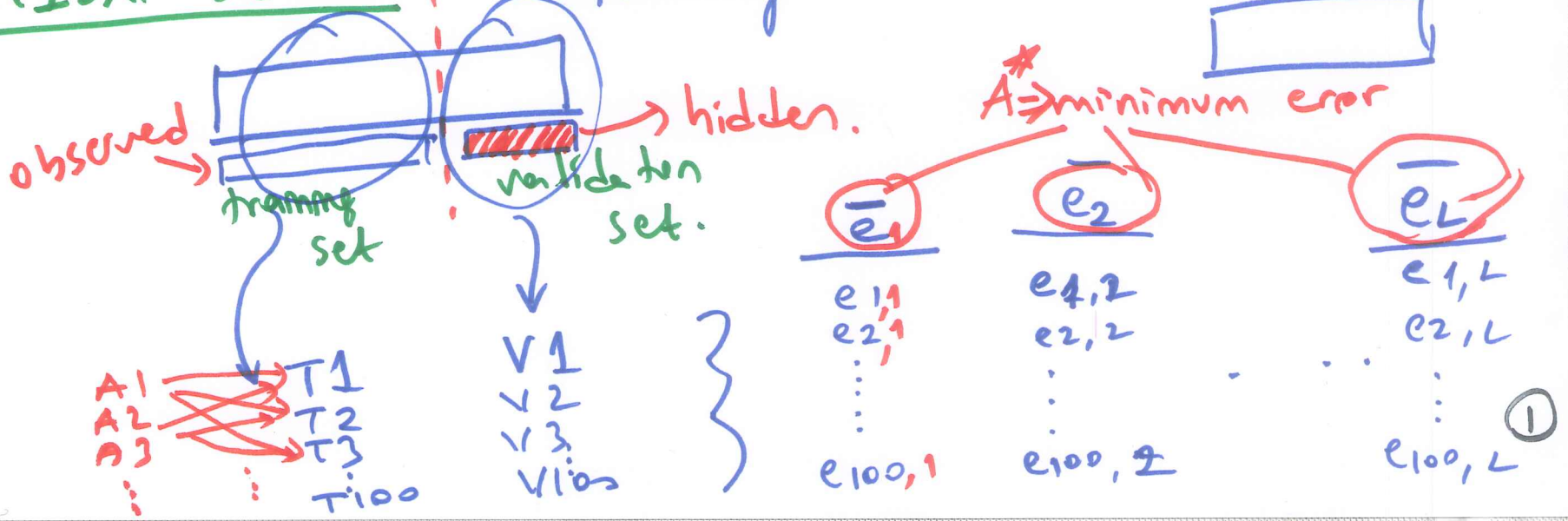
① How can we assess the expected error of a learning algorithm for a given problem?

② Given two/more algorithms, how can we say that one is better than the other(s) for a given problem?

⇒ WE CANNOT USE TRANINING SET ERRORS TO ANSWER ① and ②

training set error < ~~error set~~ test error

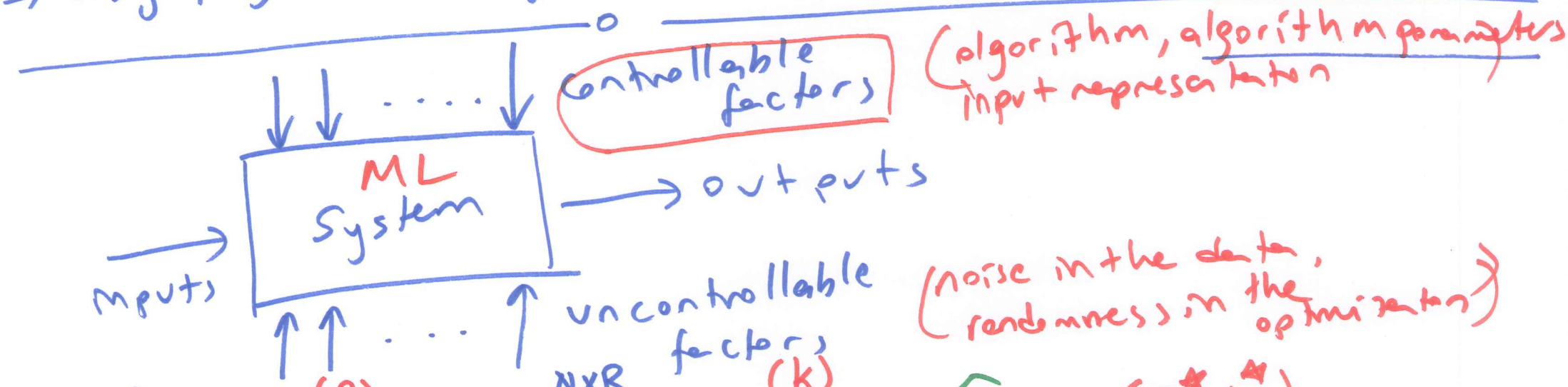
⇒ VALIDATION SETS:



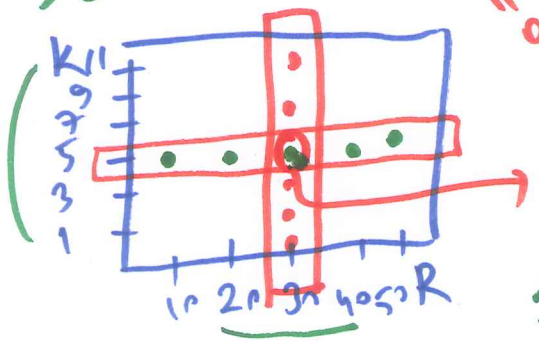
①

⇒ loss functions → minimizing FPs
→ minimizing FNs
→ minimizing the cost.

⇒ time & space complexity.
⇒ interpretability.
⇒ easy programmability.



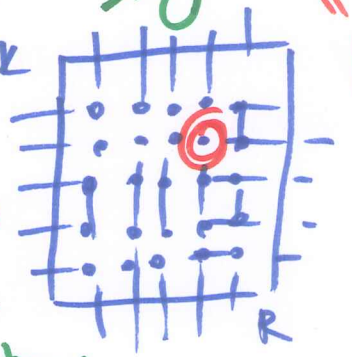
$\underset{N \times D}{X} \rightarrow PCA \xrightarrow{(R)} \underset{N \times R}{Z} \rightarrow K\text{-}NN \xrightarrow{(k)} \hat{y}$



"one factor at a time"

30 is the best given k=5

$5 + 6 - 1 = 11 - 1 = 10$ options



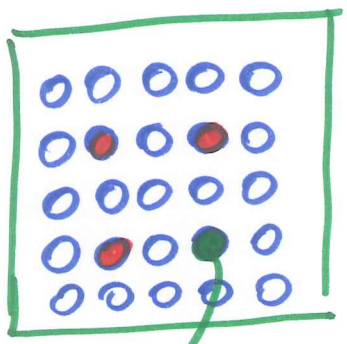
"factorial design"

(R^*, k^*)

5x6 = 30 options

(R^*, k^*) out of 30 options.

Response surface Design.



$\rightarrow (y_1)(e_1)$
$\rightarrow (y_2)(e_2)$
$\rightarrow (y_3)(e_3)$

used for the model. regression algorithm

$y_1 \rightarrow \hat{e}_1$
$y_2 \rightarrow \hat{e}_2$
$y_3 \rightarrow \hat{e}_3$
$\vdots$
$y_{25} \rightarrow \hat{e}_{25}$

GUIDELINES FOR ML EXPERIMENTS

(A) Aim of the study → evaluating a single algorithm and reporting its error
→ picking the best alg. on a specific problem.
→ picking the best alg. on several problems.

(B) Selection of the response variable] performance criteria.

(C) Choice of factors & their levels] → algorithms
→ their parameters.

(D) Choice of Experimental design] → do factorial design if possible
→ one factor at a time
→ response surface design.
⋮

(E) Run experiments → use cloud computing if available.

(F) Statistical analysis of the data results. → Alg A > Alg B
statistically significant or not?

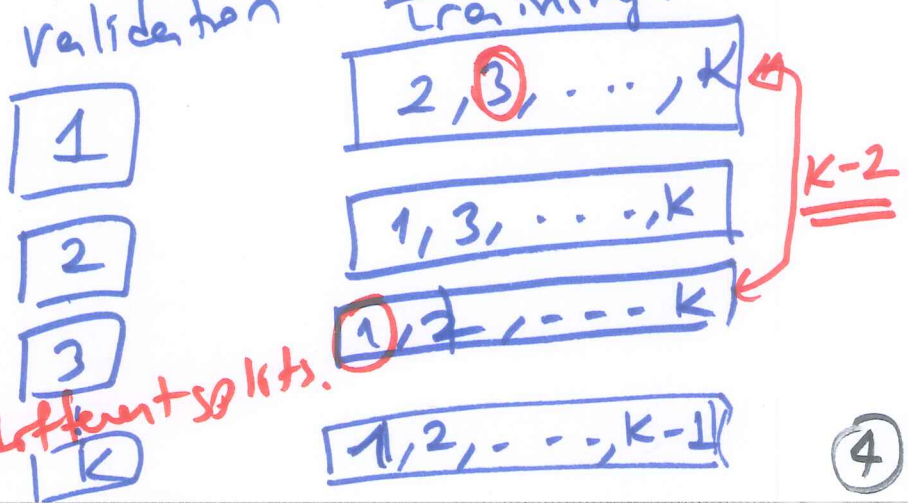
(G) Conclusions & Recommendations

CROSS-VALIDATION & RESAMPLING

100 Million data points.

K=10 ⇒ $\frac{K-2}{K-1} = \frac{8}{9}$

K-FOLD CROSS VALIDATION

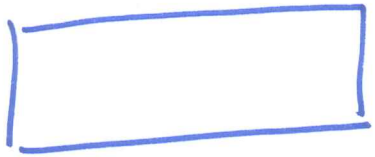


$k = N$ ⇒ LEAVE ONE-OUT CROSS-VALIDATION

40 points ⇒ 5-fold CV ⇒ 32 train, 8 validation

leave one-out ⇒ 39 train, 1 validation

5x2 CROSS VALIDATION



randomly shuffle