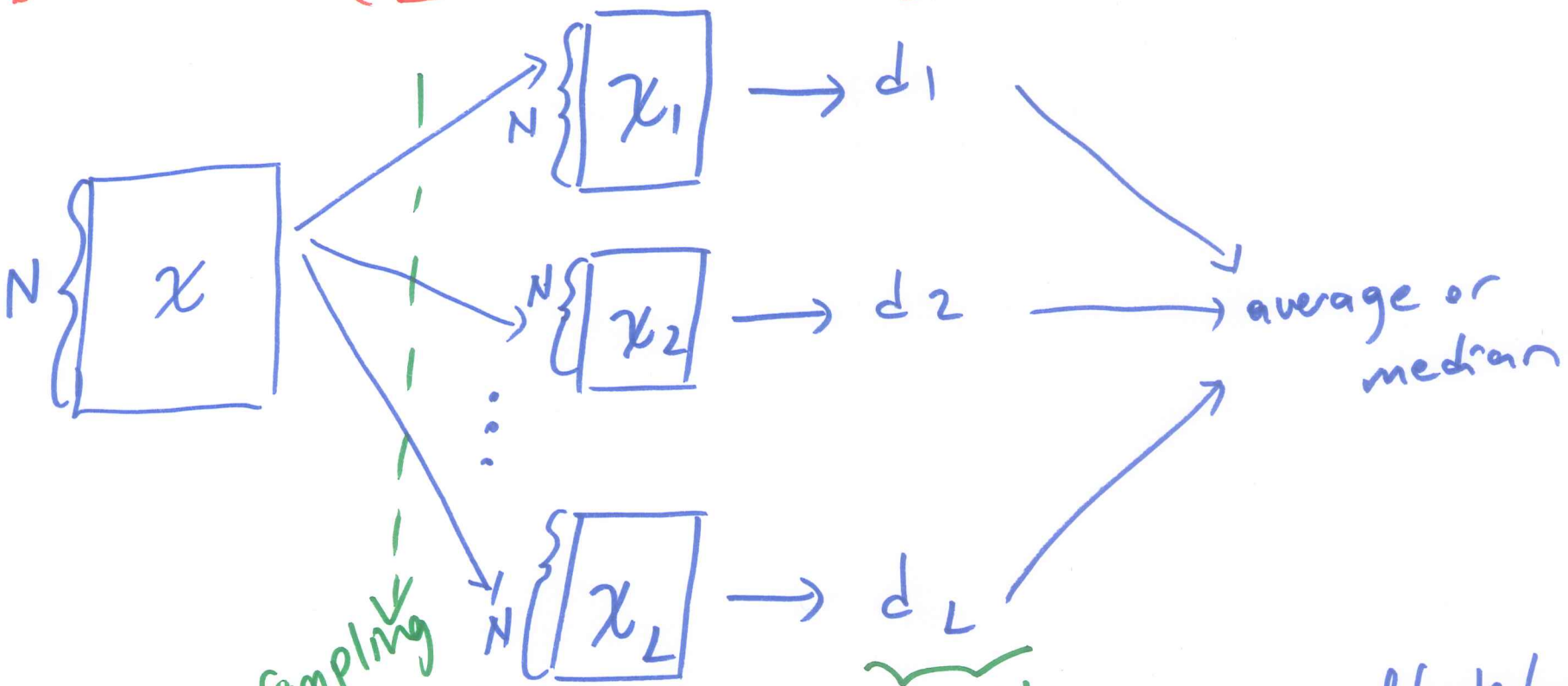
BAGGING (Bootstrap AGGregation)

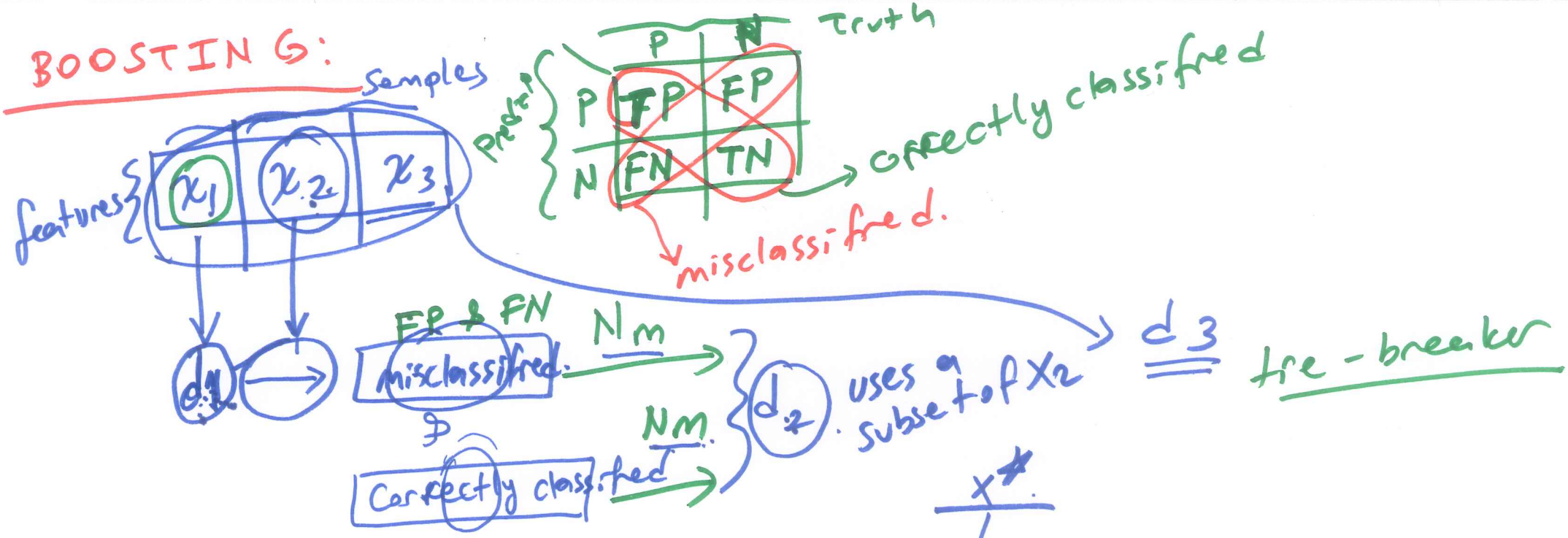


unstable
decision tree
MLP

stable
k-NN

If N is very large, N' < N data points should be picked so training sets would become different

BOOSTING:

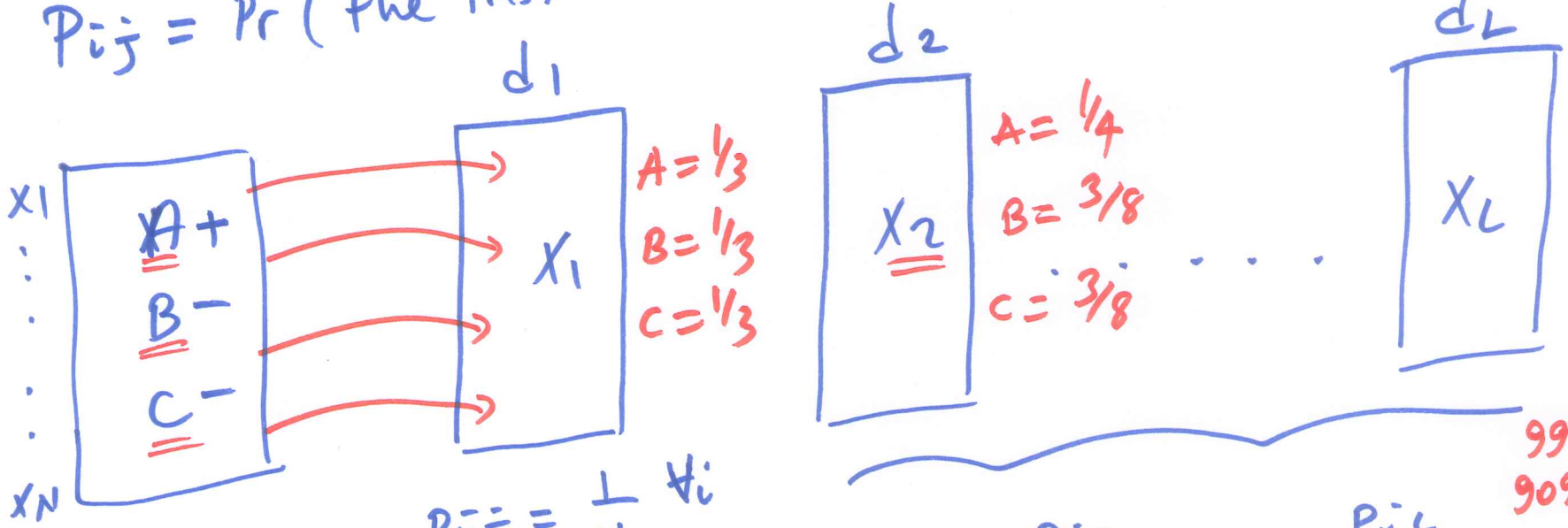


⇒ if they agree, no problem use the predicted label.

⇒ if they do not agree, use $d_3(x^*)$

AdaBoost: modify the probabilities of drawing instances as a function of the error.

P_{ij} = Pr(the instance x_i is selected by classifier d_j)



$P_{ij} = \frac{1}{N} \quad \forall i$

$P_{i2}, P_{i3}, \ldots, P_{iL}$

99% → log(99)
90% → log(9)

ϵ_j → error rate of classifier j.

$\frac{\epsilon_j}{1-\epsilon_j}$

0.10 0.01 0.90 0.99

x^*

⇓ decreased for correctly classified data points.

$d_1(x^*)$ $d_2(x^*)$ . . . . $d_L(x^*)$

w_1 w_2 w_L

$w_j = ?$

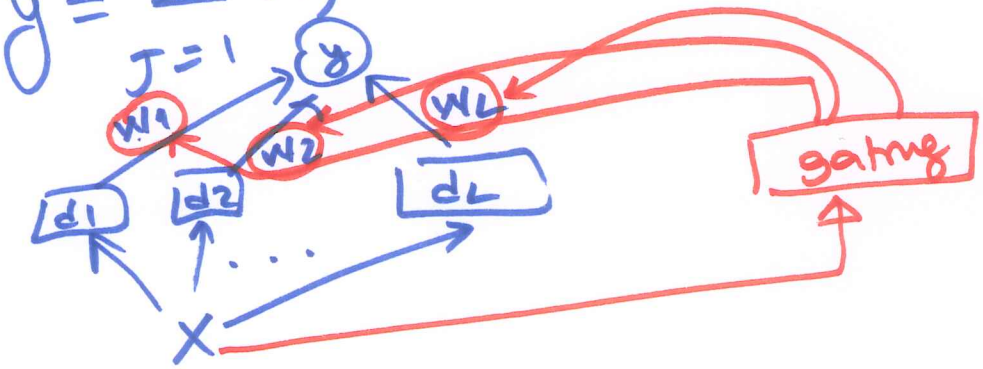
$w_j = \log\left(\frac{1}{\beta_j}\right)$

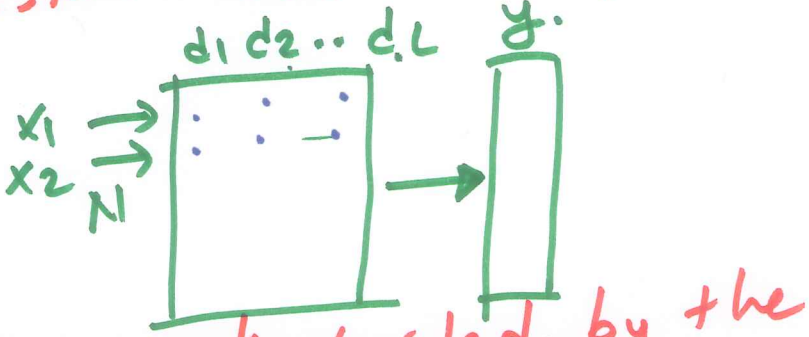
$d(x^*) = w_1 \cdot d_1(x^*) + w_2 \cdot d_2(x^*) + \ldots + w_L d_L(x^*)$

Mixture of Experts (MoE)

Voting $\Rightarrow y = \sum_{j=1}^{L} w_j d_j(x^*)$ → constant over the input space.

$y = \sum_{j=1}^{L} w_j(x^*) \cdot d_j(x^*)$





will be decided by the gating model.

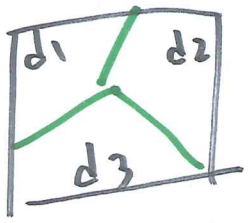
Competitive

→ $w_1, w_2, \ldots w_L$ are usually sparse

→ one or some of them are nonzero.

Softmax.

↳ $v_{ij} = \dfrac{\exp(v_j^T \cdot x + v_{j0})}{\sum_{k=1}^{L} \exp(v_k^T \cdot x + v_{k0})}$

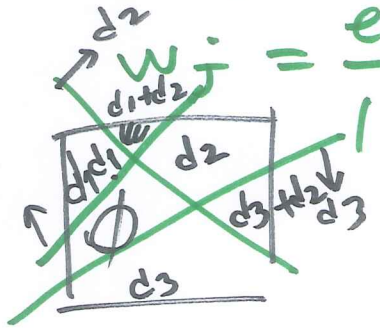


Cooperative

→ $w_1, w_2, \ldots w_L$ are assumed to be independent

→ sigmoid

$w_j = \dfrac{\exp(-[v_j^T \cdot x + v_{j0}])}{1 + \exp(-[v_j^T \cdot x + v_{j0}])}$



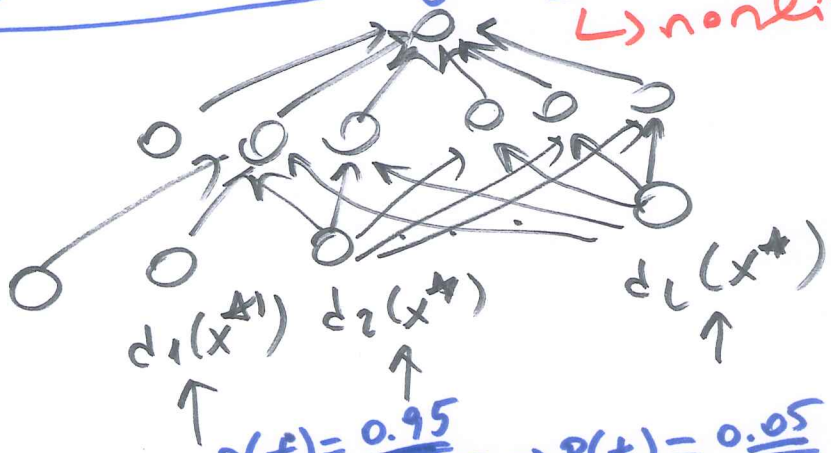
Stacked Generalization

Voting: $\sum_{j=1}^{L} w_j d_j(x^*)$

MoE: $\sum_{j=1}^{L} w_j(x^*) d_j(x^*)$

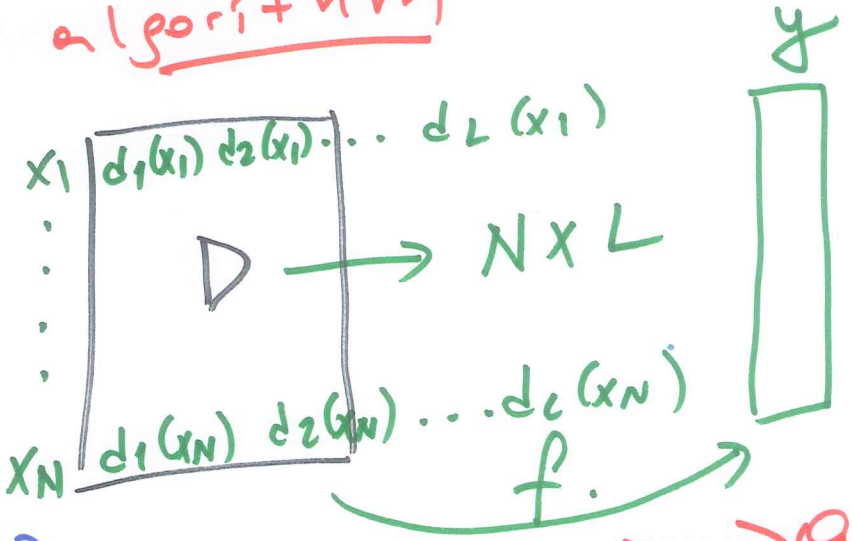
Stacked Generalization $y = f\left(d_1(x^*), d_2(x^*), \ldots d_L(x^*) \mid \Phi\right)$

↳ nonlinear algorithm



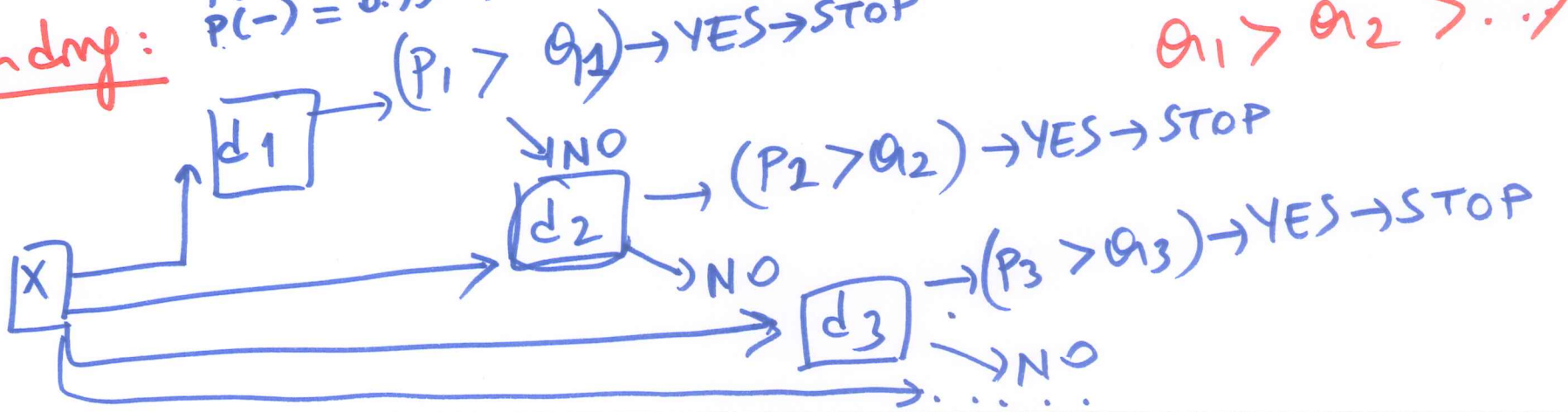
$P(f) = \frac{0.95}{0.95} \Rightarrow P(+) = 0.05$

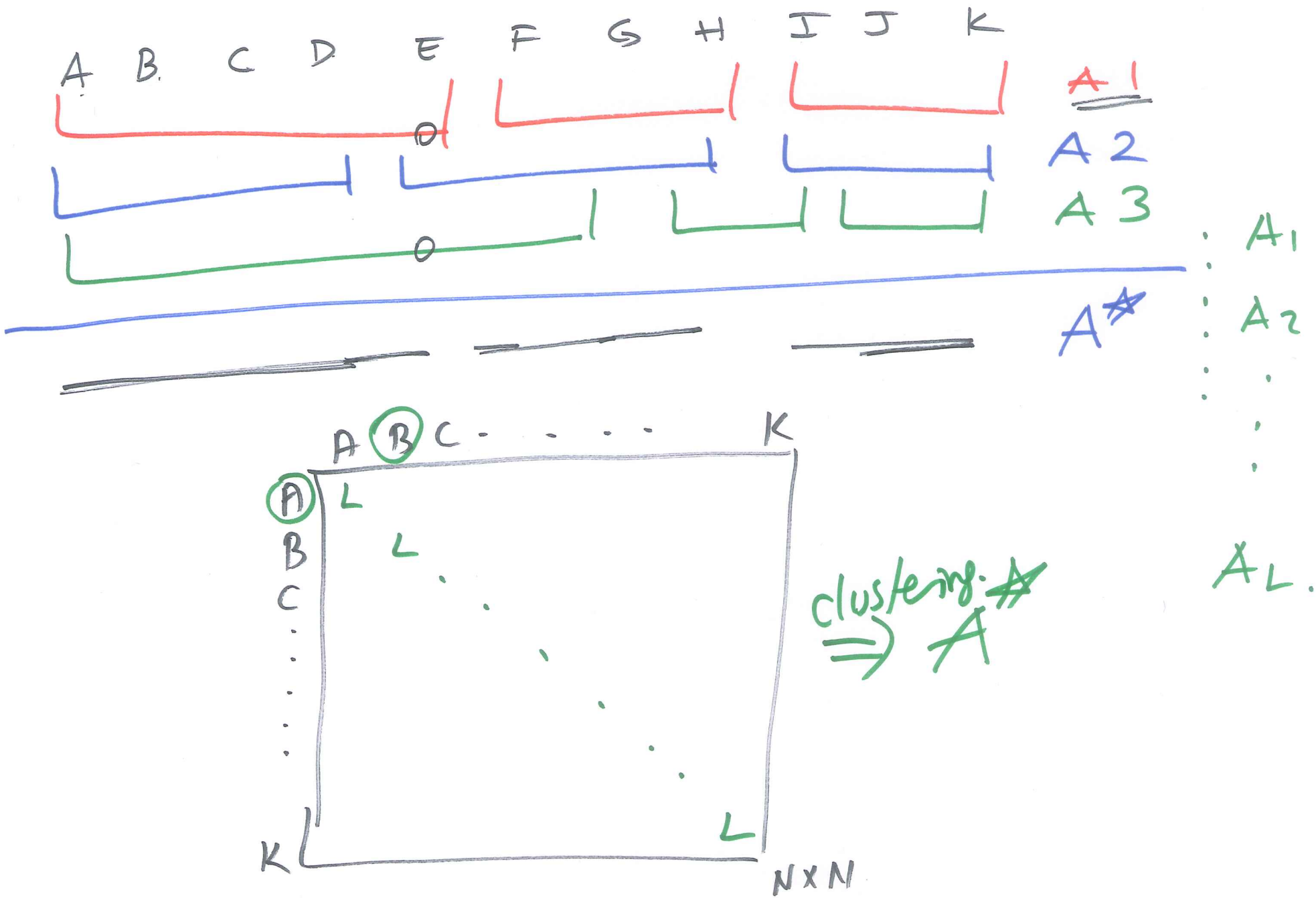
$P(-) = 0.95$



$\theta_1 > \theta_2 > \ldots > \theta_L$

Cascading:




A B C D E F G H I J K
A 1
A 2
A 3
A*
A1
A2
AL.
A B C K
A
B
C
K
L
L
L
N x N
clustering
A*