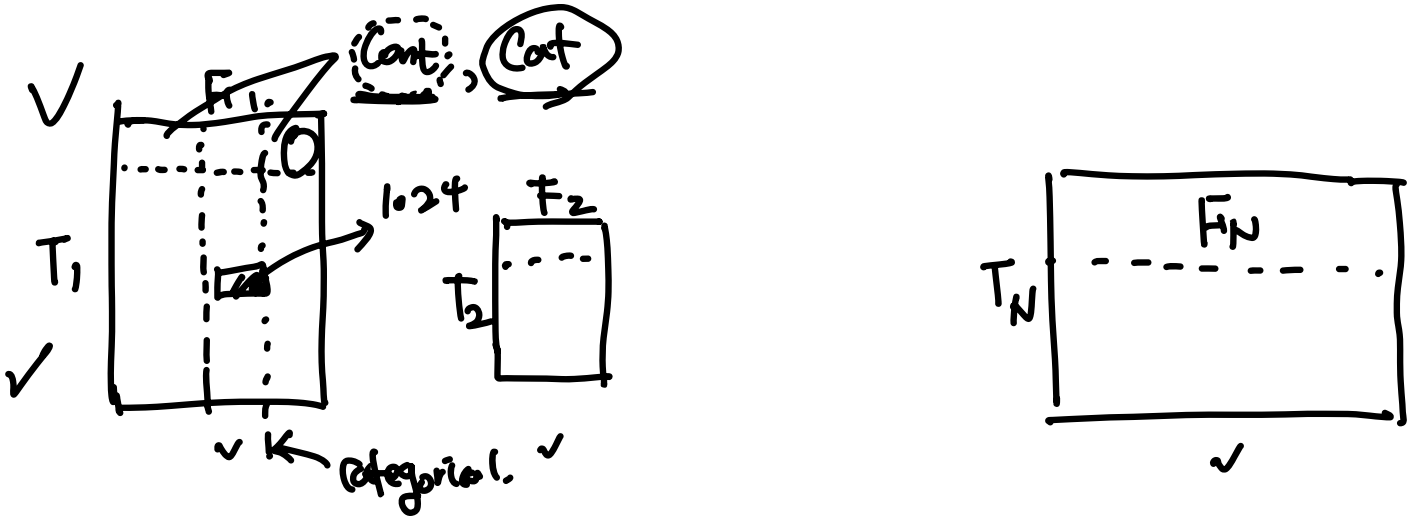
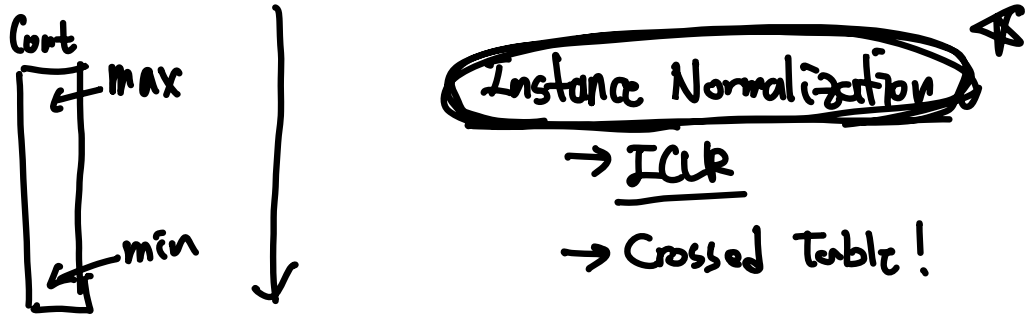
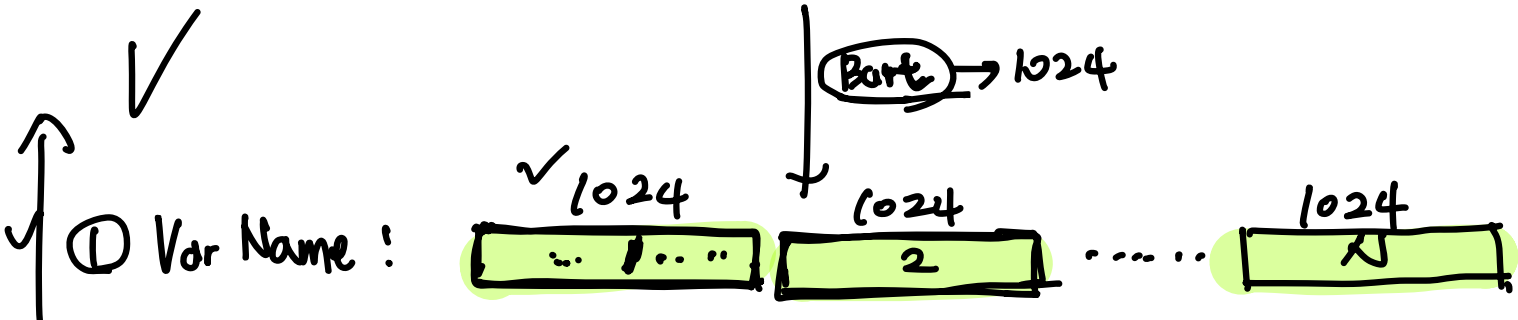




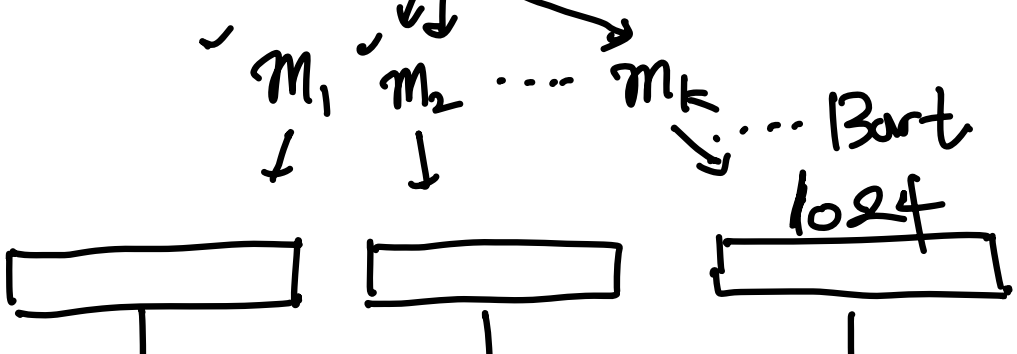
① Feature name : T_1 T_2 ... T_N.

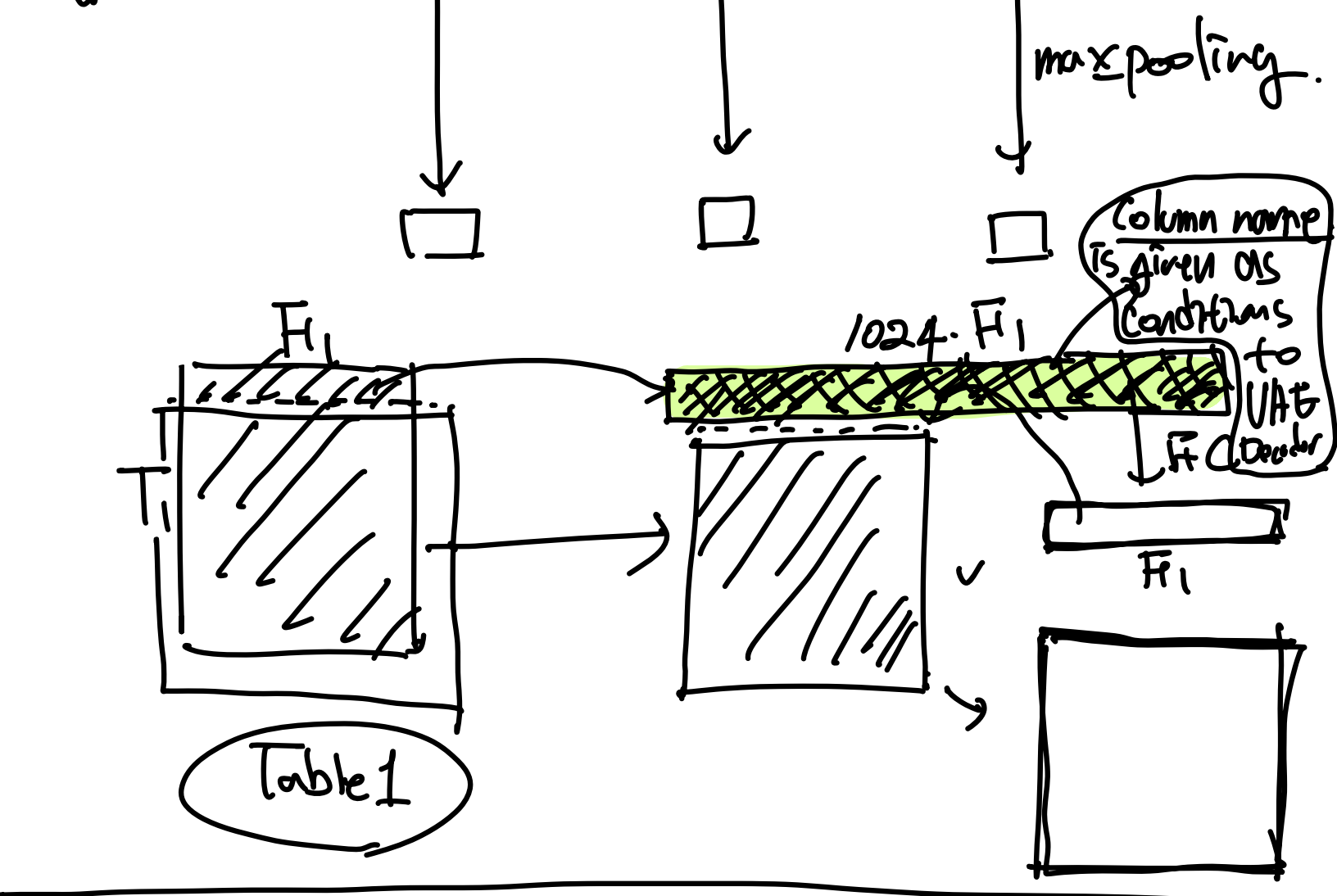
↓ ↓ ↓

F_1 F_2 F_N

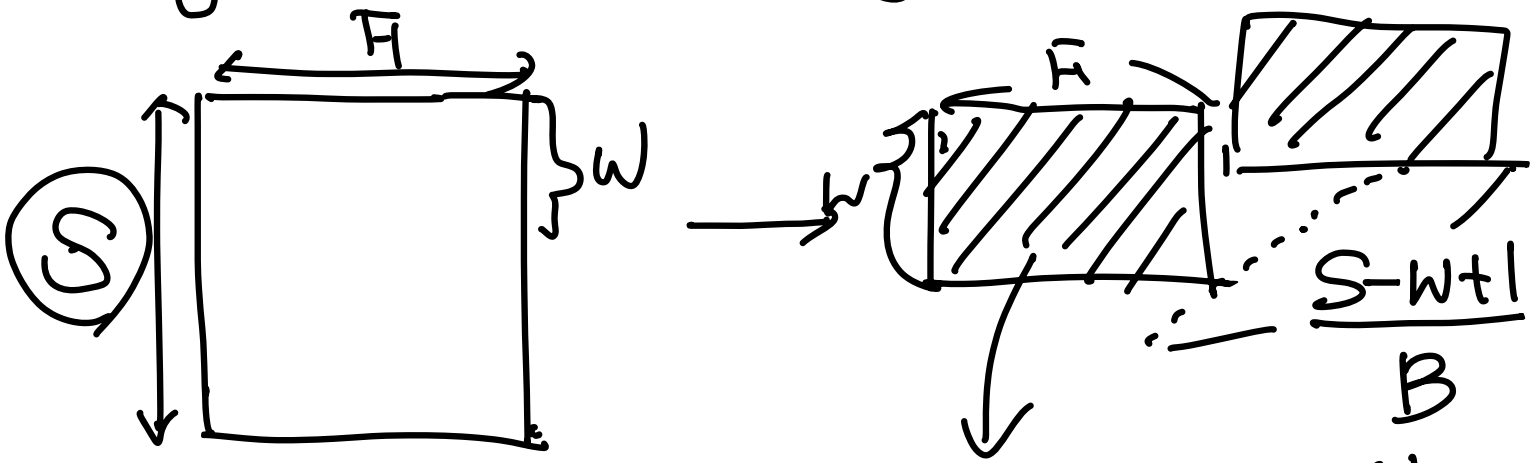


② Categorical Var : if k categorical variables

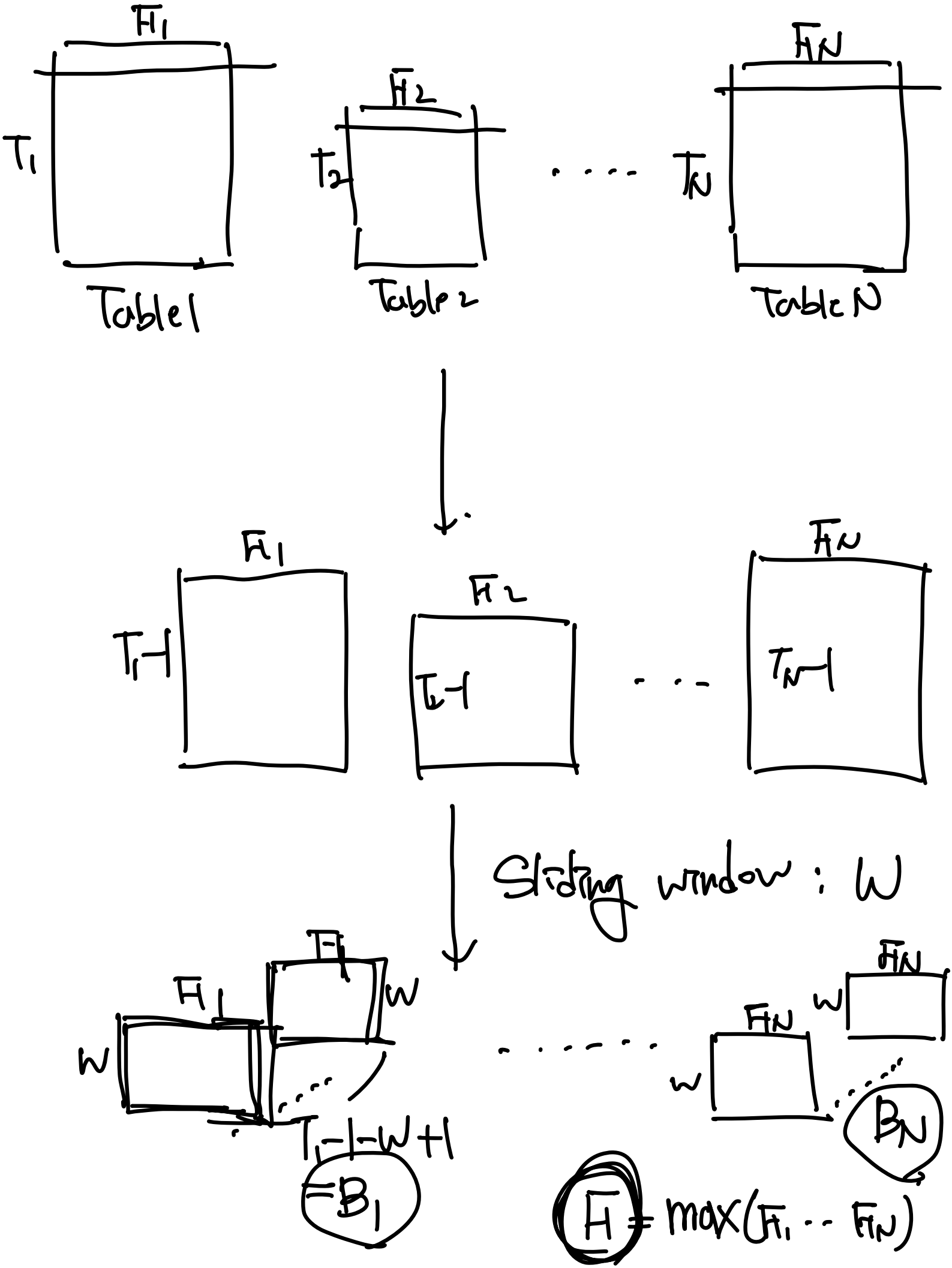




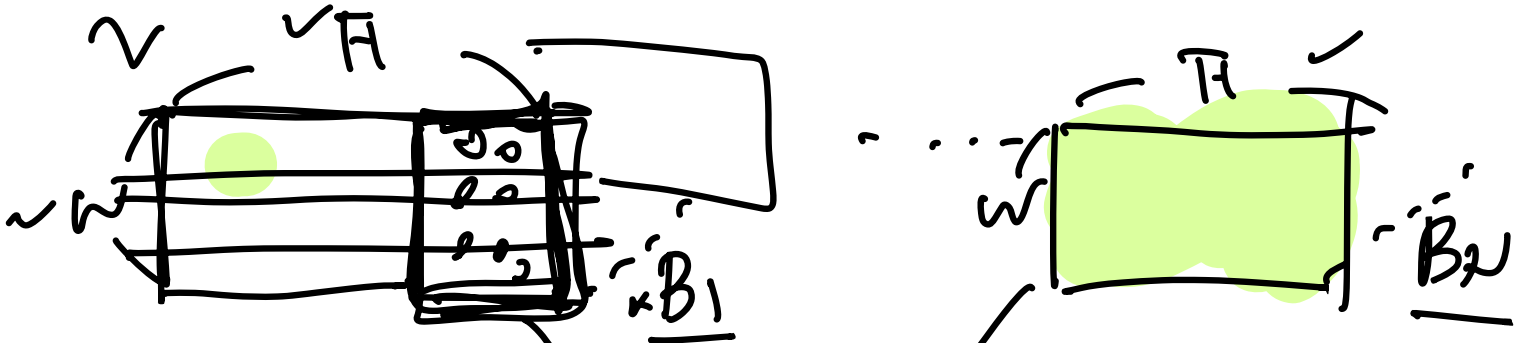
Single Tabular Data Learning:



We think each of the data point is same distribution.


F1
T1
Table1
F2
T2
Table2
FN
TN
TableN
F1
T1-1
F2
T2-1
FN
TN-1
Sliding window : W
F1
W
W
T1-1-W+1
=B1
FN
W
BN
H = max(F1 .. FN)

Instance Normalization + Patching.



Preprocessing.

VAE:

Encoder from TimeAutoDiff.

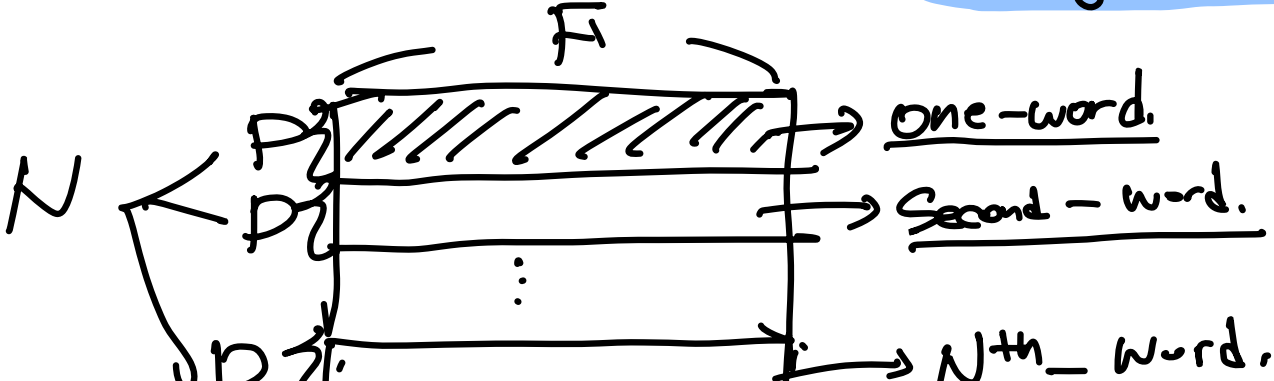
Koopman:
Decompose: Static + Changing over time.

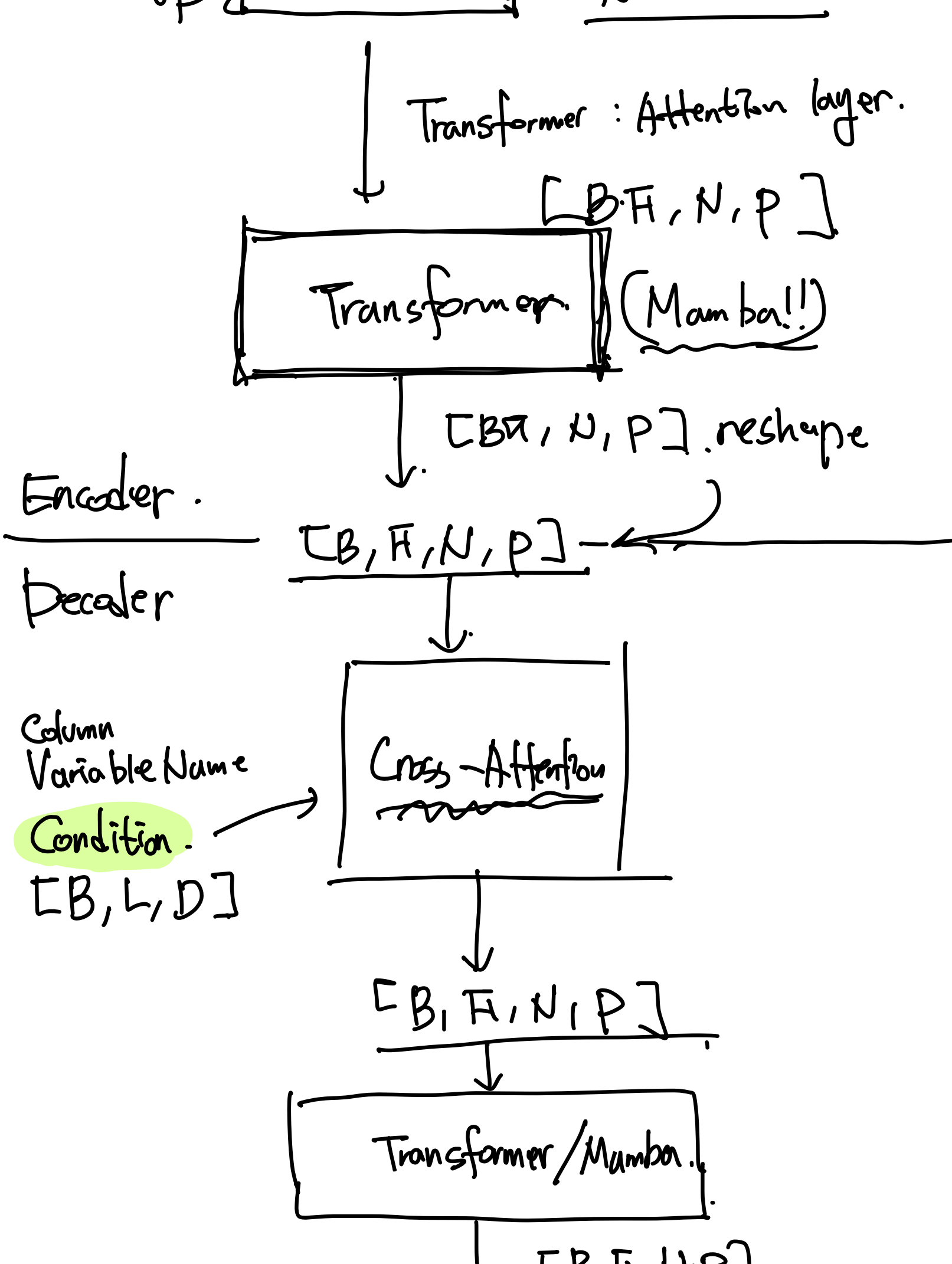
# of Patches

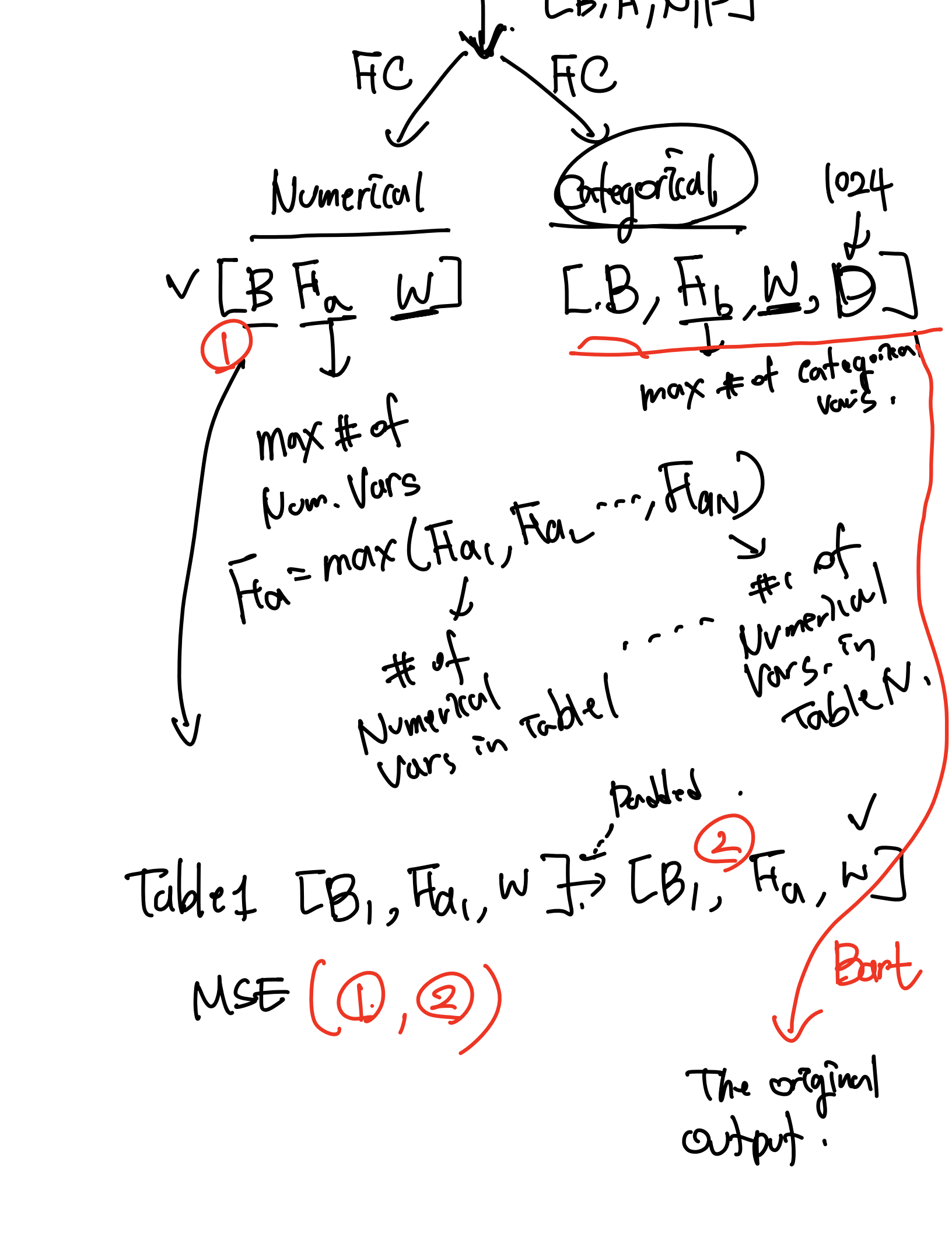
Patch Size

[B1, F, W] ⟶ [B1, F, N, P]

Patching - Embedding.




Transformer : Attention layer.
[B·F, N, P]
Transformer
(Mamba!!)
[BF, N, P].reshape
Encoder.
Decoder
[B, F, N, P]
Column
Variable Name
Condition.
[B, L, D]
Cross-Attention
[B, F, N, P]
Transformer/Mamba.


FC
FC
Numerical
Categorical
1024
[B Fa W]
[B, Fb, W, D]
max # of Num. Vars
max # of categorical vars.
Fa = max(Fa1, Fa2 ..., FaN)
# of Numerical Vars in Table1
# of Numerical Vars. in TableN.
Padded.
Table1 [B1, Fa1, W] → [B1, Fa, W]
MSE (1, 2)
Bart
The original output.