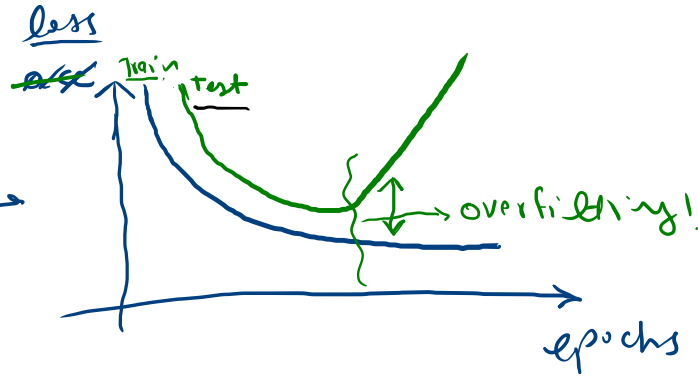
Neural network enhancement

overfitting →
- Dropout
- data augmentation / more data
- Batch normalization

underfitting →
- Complex model / deeper model (more neurons) (more layer)
- Data augmentation / more data
- Hp tuning!

- Train - acc
- Test - acc

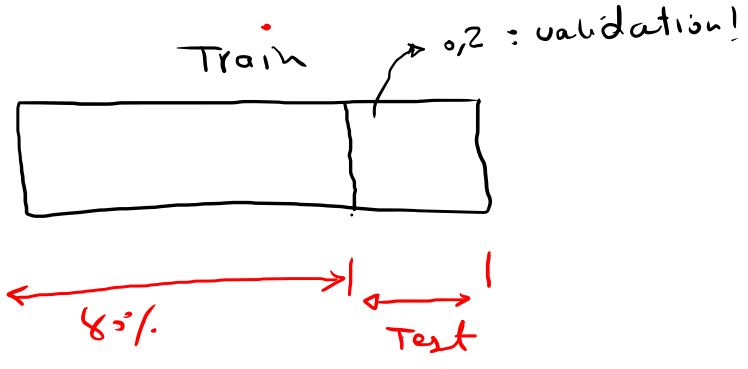


overfitting → Train acc ↑ / Test acc ↓

- Train-data → acc, loss / epoch
  - done
- Test-dat → acc-loss

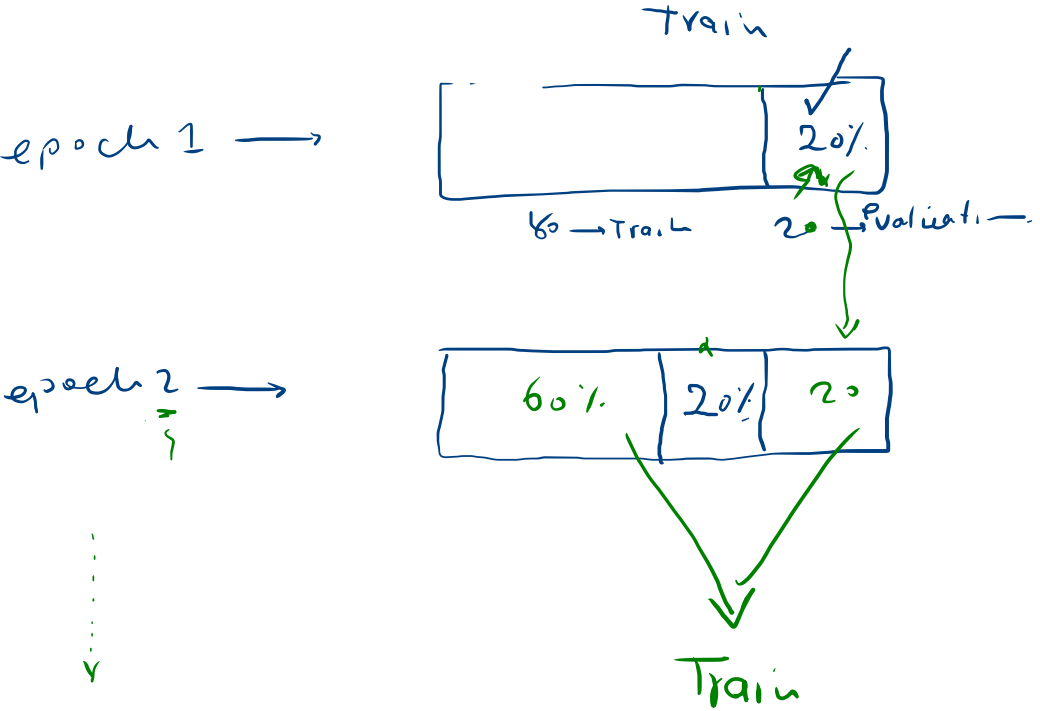
dev ..

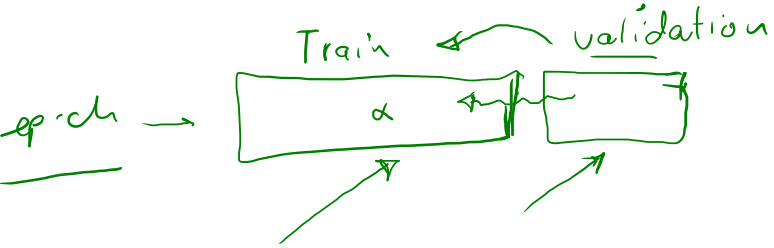
→ validation-set



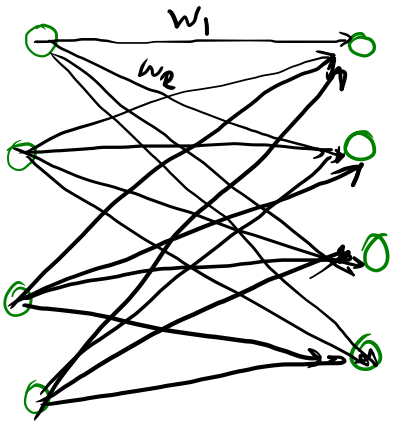
Train ace → /60 → underfitting *

Test ace → /20 → overfitting


Train
epoch 1
20%
60% → Train
20% → Validati
epoch 2
60%
20%
20
Train


Train
validation
epoch
α

Dropout



$4 \times 4 = 16$

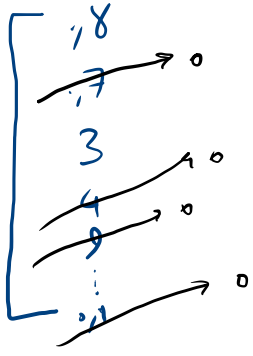
$4 \longrightarrow 0$

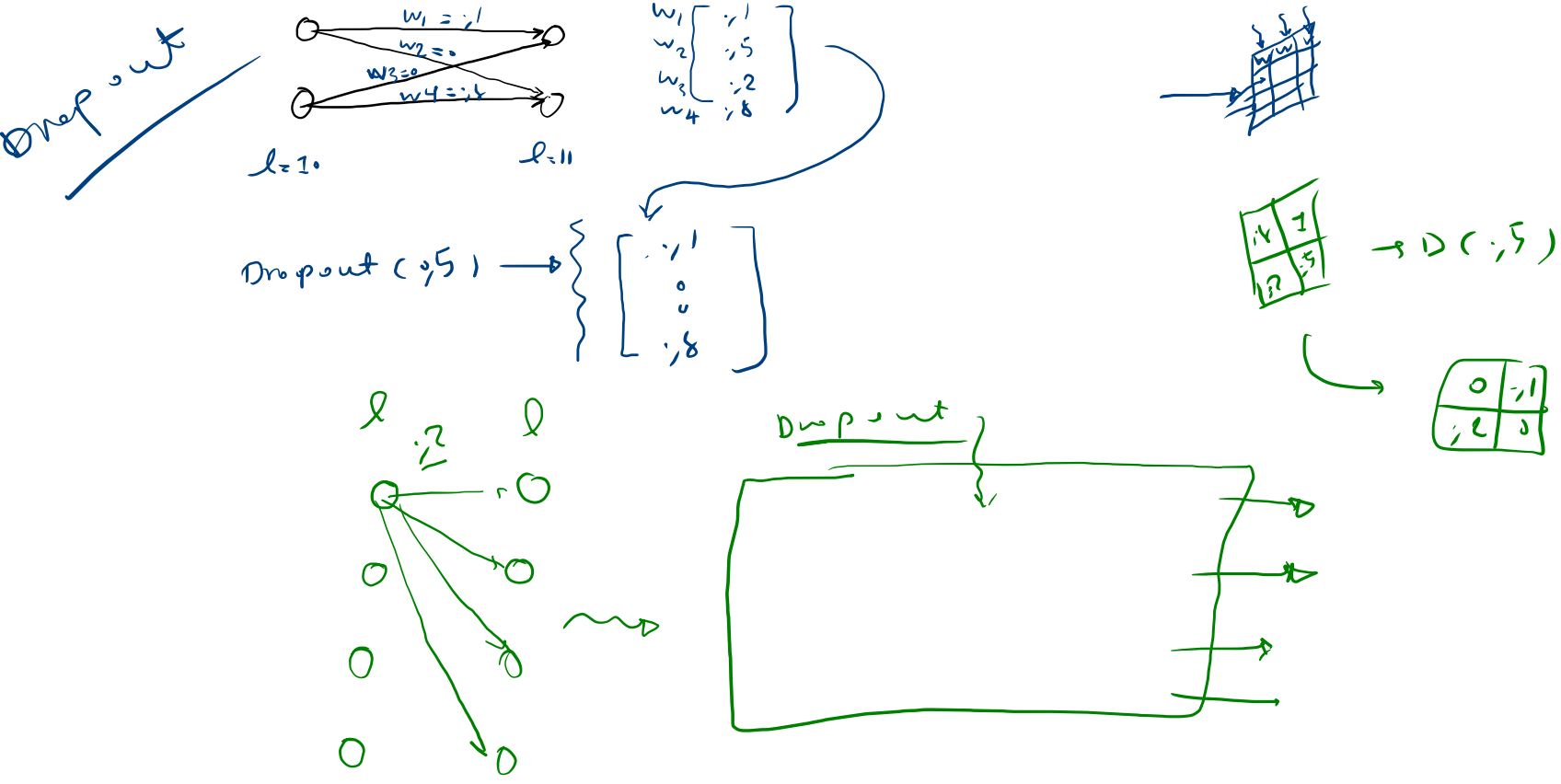
Dropout (.25)

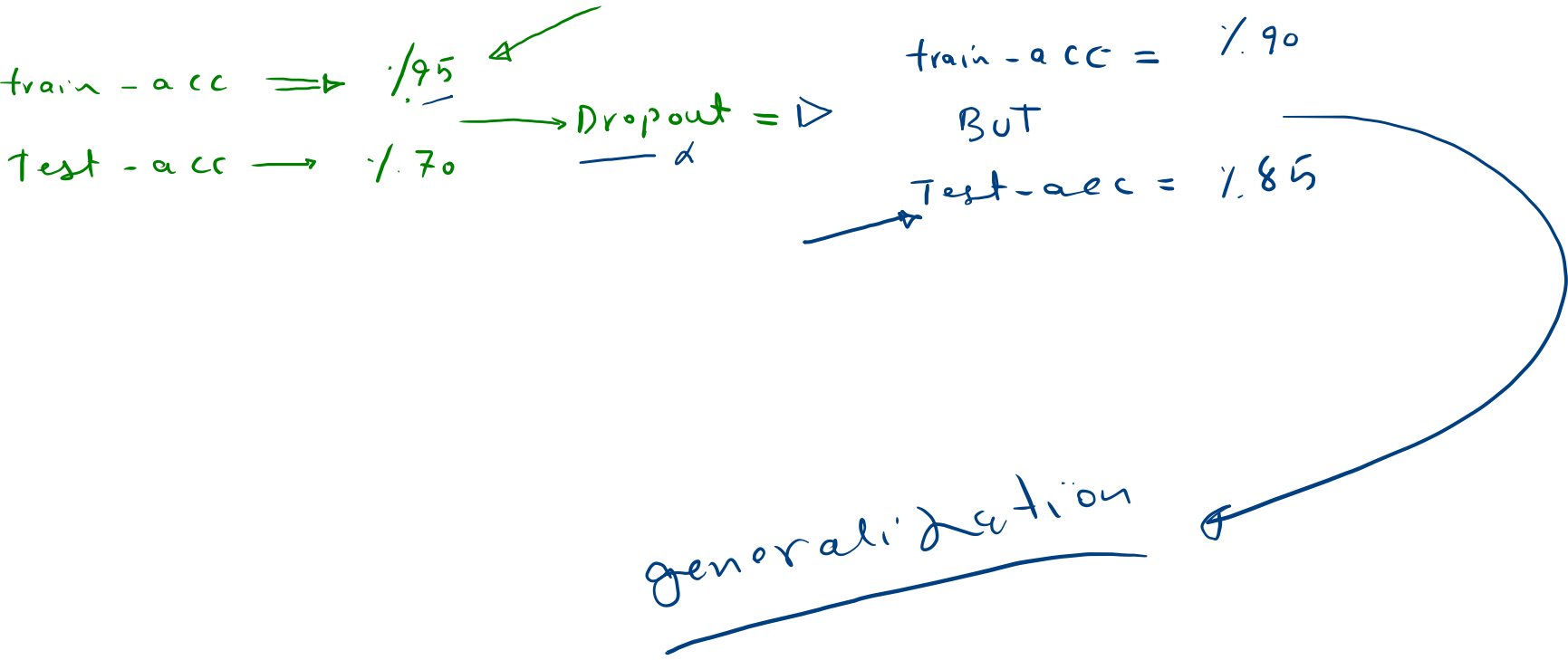
.25 ← 25% → random → 0

layer = n layer : n+1

w




Dropout
Dropout (0,5)
Dropout


train - acc ⟹ ٪95
test - acc ⟶ ٪70
Dropout ⟹
train - acc = ٪90
But
Test - acc = ٪85
generalization

Batch - Size = 3

Batch Normalization

$net^{(1)} = w_1 x_1 + w_2 x_2 + \dots$

$net^{(2)} = w_1 x \dots$

$net^{(3)} = w_1 \dots$

w $\quad$ b

$O_1 = \gamma\, \sigma(net^{(1)}) + \beta$

learnable parameters

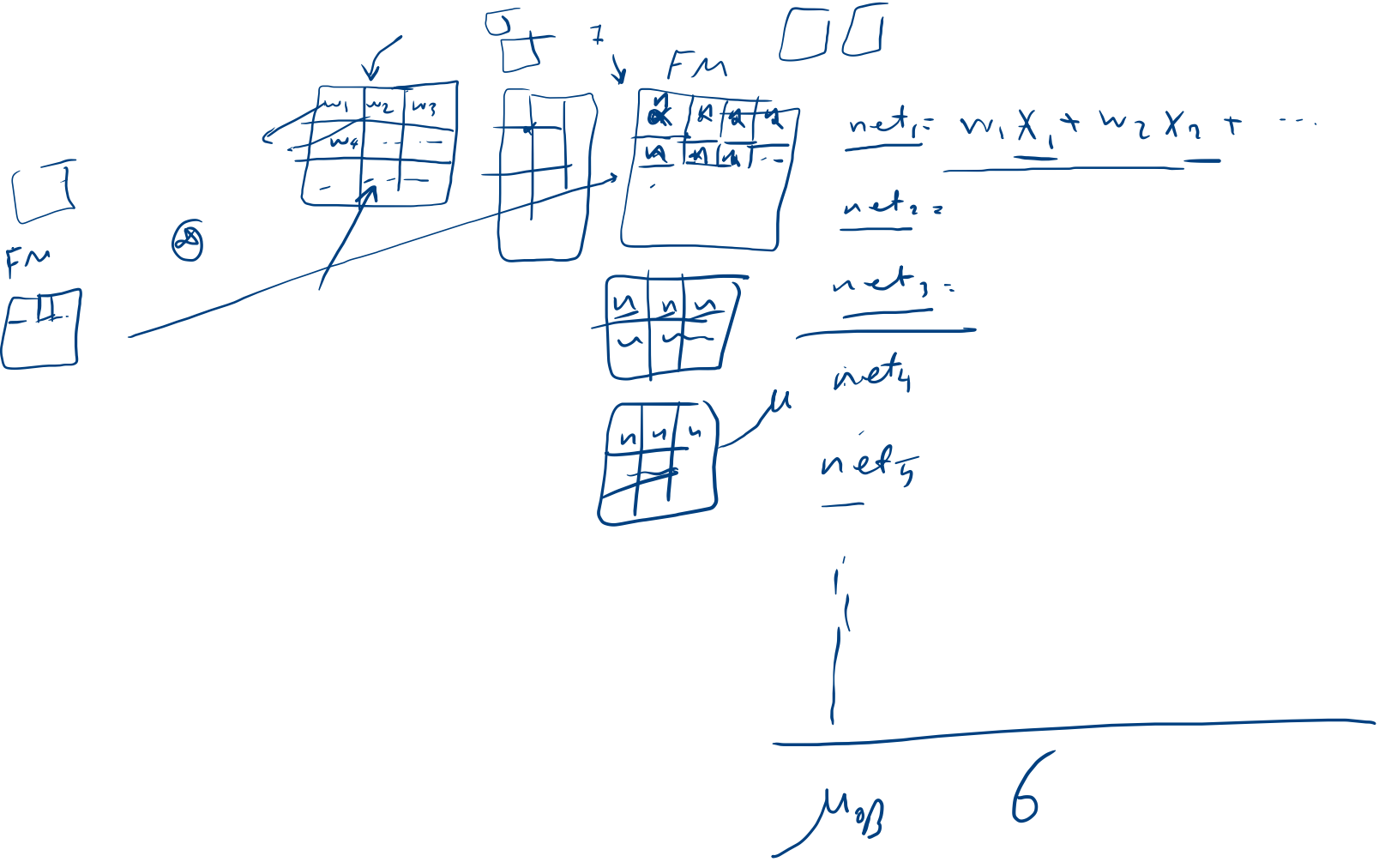
~~$\sigma(net)$~~

$$\mu_B = \frac{net^{(1)} + net^{(2)} + net^{(3)}}{3}$$

SD

$$\sigma_B = \sqrt{\frac{1}{3}\left[(net^{(1)} - \mu_B)^2 + (net^{(2)} - \mu_B)^2 + (net^{(3)} - \mu_B)^2\right]}$$

$$\hat{net}^{(1)} = \frac{net^{(1)} - \mu_B}{\sigma_B} \qquad \hat{net}^{(2)} = \frac{net^{(2)} - \mu_B}{\sigma_B} \qquad \hat{net}^{(3)} = \frac{net^{(3)} - \mu_B}{\sigma_B}$$


FM
FM
w_1 w_2 w_3
w_4
$net_1 = w_1 x_1 + w_2 x_2 + \cdots$
$net_2 =$
$net_3 =$
net_4
net_5
μ
$\mu_{0.3}$
6


End