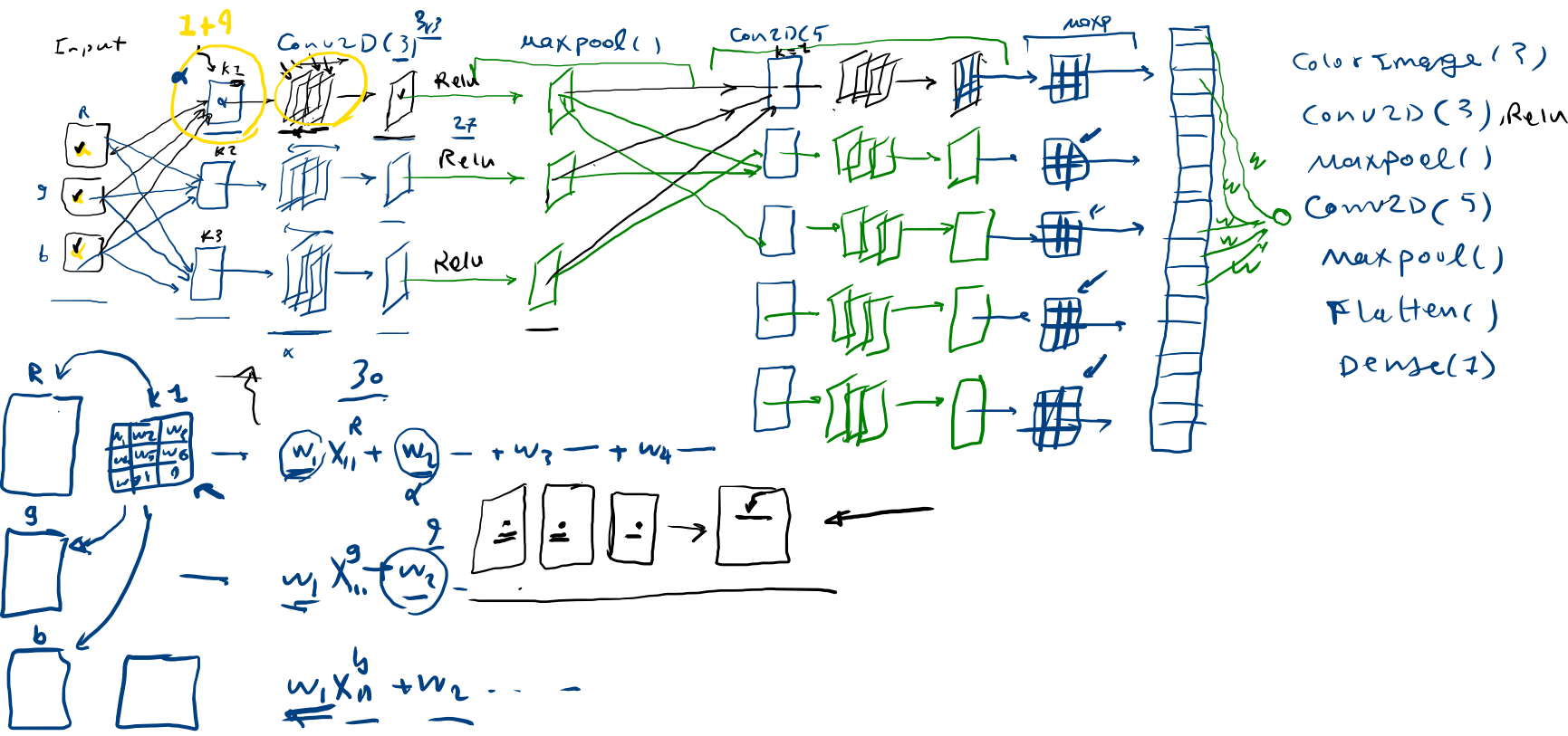
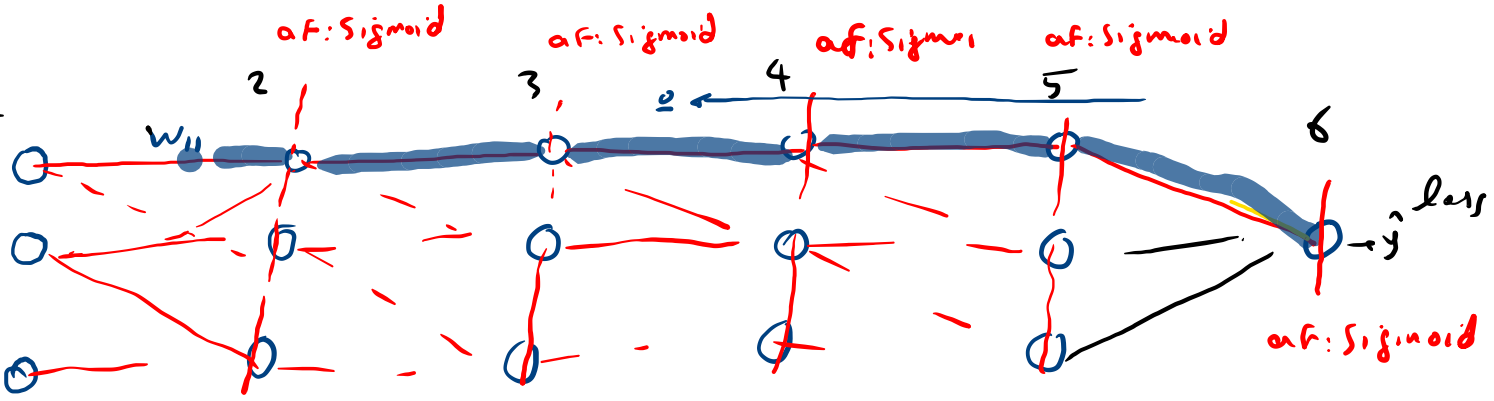
Neural Network

Improvement


Input
1+9
Conv2D(3)
3x3
maxpool()
Conv2D(5
maxp
R
g
b
K1
K2
K3
Relu
27
Relu
Relu
Color Image(3)
Conv2D(3), Relu
maxpool()
Conv2D(5)
maxpool()
Flatten()
Dense(1)
30
R
K1
g
b
$w_1 X_{11}^R + w_2 - + w_3 - + w_4 -$
$w_1 X_{11}^g + w_2 -$
$w_1 X_{11}^b + w_2 \ldots -$

5 2006

Sig / tanh



$0.1 \times .2 \times 0.1 \times .15 \times .2 \simeq 0$

$w_{11}^{new} = w_{11}^{old} - \alpha \frac{\partial L}{\partial w_{11}}$

$$\frac{\partial L}{\partial w_{11}} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial o_6} \frac{\partial o_6}{\partial net_6} \frac{\partial net_6}{\partial o_5} \frac{\partial o_5}{\partial net_5} \frac{\partial net_5}{\partial o_4} \frac{\partial o_4}{\partial net_4} \frac{\partial net_4}{\partial o_3} \frac{\partial o_3}{\partial net_3} \frac{\partial net_3}{\partial o_2} \frac{\partial o_2}{\partial net_2} \frac{\partial net_2}{\partial w_{11}}$$

0.1 · 0.2 · 0.1 · 0.5 · 0.2 · 0.1 · 0.15 · 0.2

اگر ما از Sigmoid و tanh که مشتقات کمتر از ۱ است!

بهتر به منظور تابع نمایی از Hidden Layers استفاده کنیم

احتمال صفر شدن مقدار گرادیان ($\frac{\partial L}{\partial w}$) بسیار زیاد است!

Vanishing Gradient

وزن‌های اولیه (لایه‌های اول)

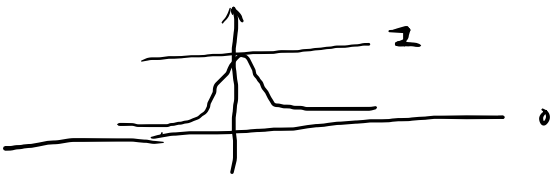
به دست نمی‌شوند!

$w_{11\text{-new}} = w_{11\text{-old}} - 0$

loss هرگز چشم‌گیری نخواهد داشت!

Jeel, ⇝ Relu

tanh
Sigmoid
Softmax



∝ exploding gradient ⇝ RNN,

Train accuracy

Test accuracy

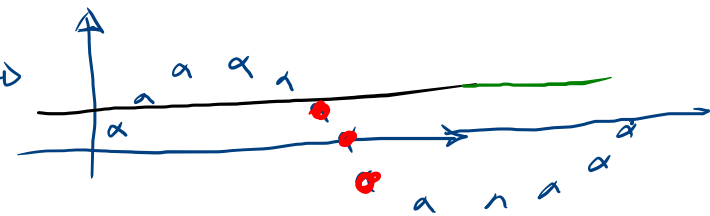
Train

دقت مدل روی داده‌های

Train-ace ⟹ low 10 % ⟶ یه درد نمی‌خوره!

high bias , underfitting!

مدل ما جواب دیتاها را یاد نگرفته است!



✓ Tools

1. complex model
   - more layers
   - more neurons

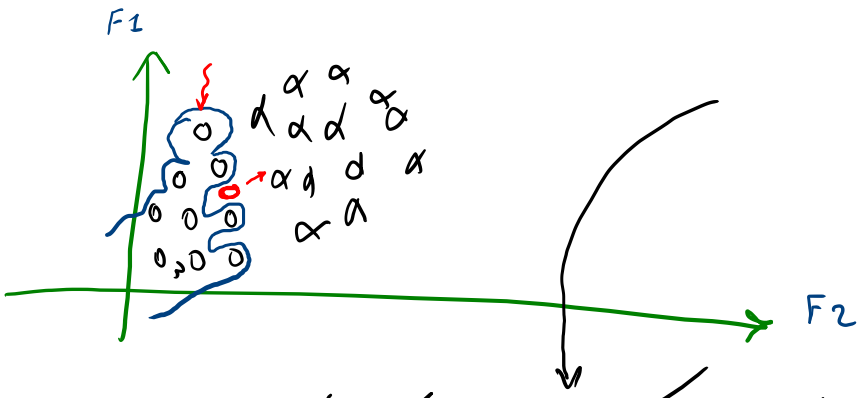
2. more data
   - bigger dataset
   - Data augmentation!

3. HP Tuning
   - α learning rate.
   - α batch size

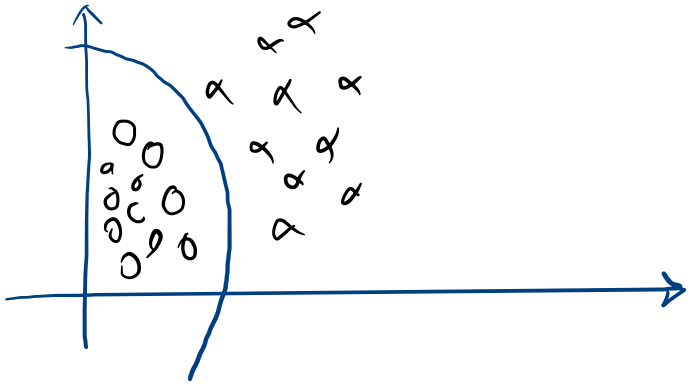
4. Suitable loss
   - reg: MSE
   - Class: cross-entropy

train-acc → high %48 → overfitting

test-acc → low %60 → high variance



نشانه این است که مدل ما به جای یادگیری مینیمم های اصلی نویزها را یاد گرفته است!



α ورزها مقادیر پایین خواهند داشت!

10,000,000

Ref,

- Regularization((1,12) ← Next
- Dropout → Next
- early stop → Next
- Data augmentation. → Next + more data
- Batch normalization → Next

90% → adam

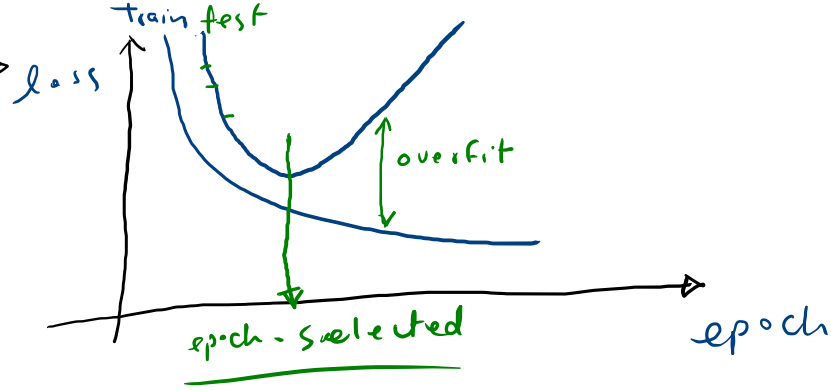
SGD

RMprop

Adagrad

⋮

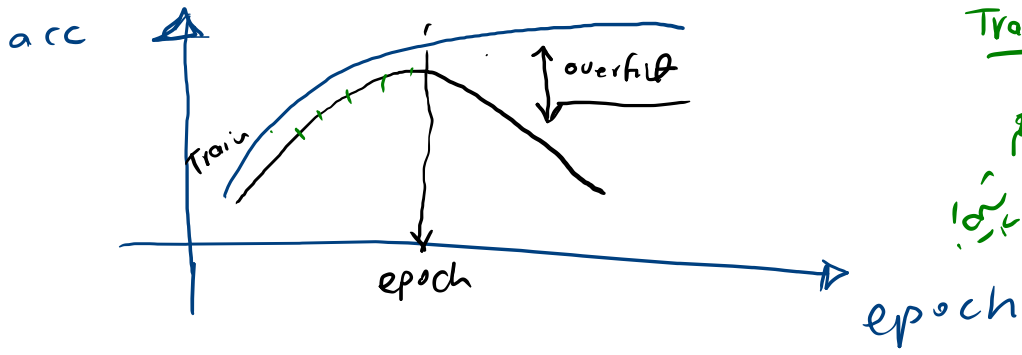
؟ } epoch تعداد



نیخ

Train
- ep1 → Train-acc
- ep2 → T-acc
- ep3 → T-acc

Test
- Test-acc

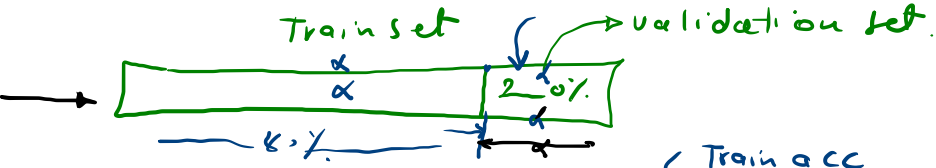


Train نیاز داریم که در حین

ما روی دیتای Test هم

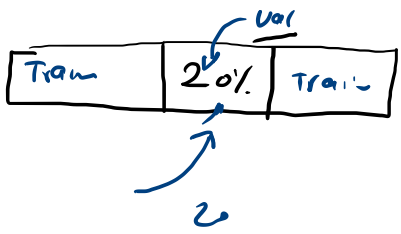
دقت و loss را داشته باشیم!

acc



ep1 → 80% — Test → Train acc, Train loss, val-acc, val-loss

ep2 →



ep3

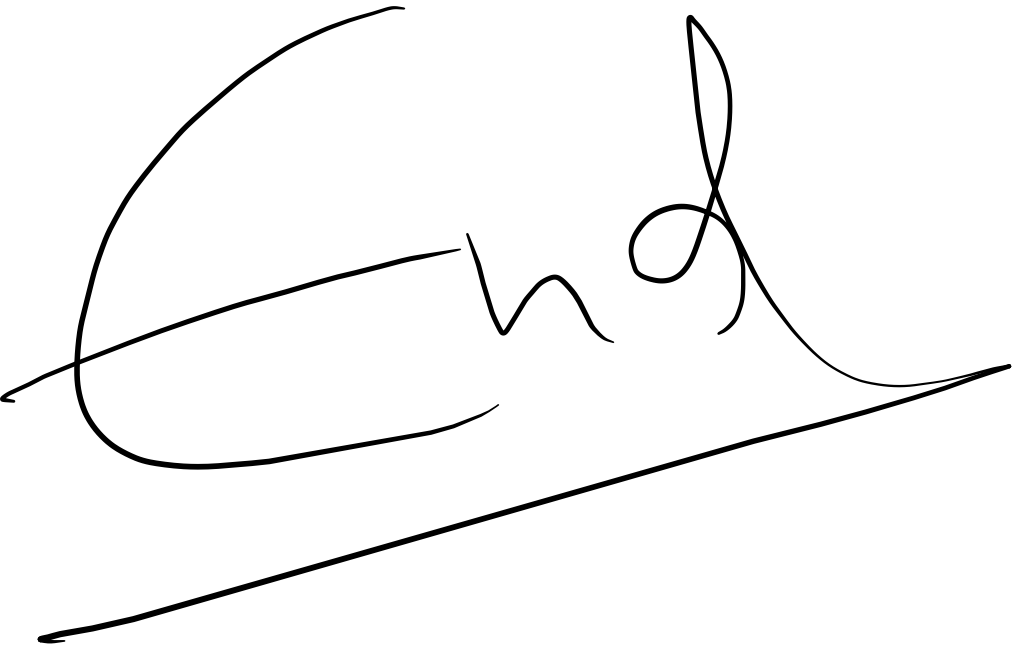
ep4 20

ep5 → 20%

validation

می‌تواند غیر از داده‌ی Train باشد؟

ok



train-ace → high

test-ace → high

→ مدل ما قابل اعتماد است!

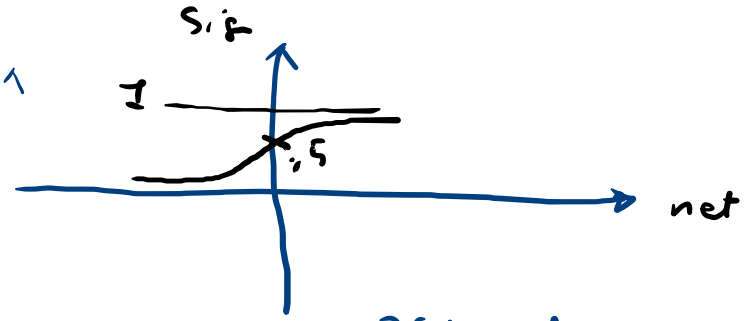
اما

تا زمانی که در شرایط واقعی تست نکرده‌ایم، نمی‌توان قضاوت قطعی

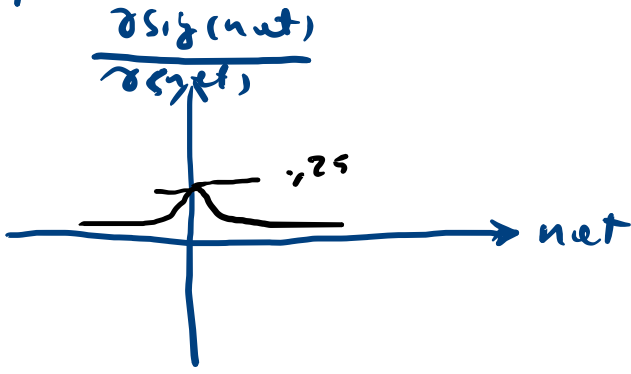
در مورد مدل NN حرفمان داشته باشیم!

→ More data

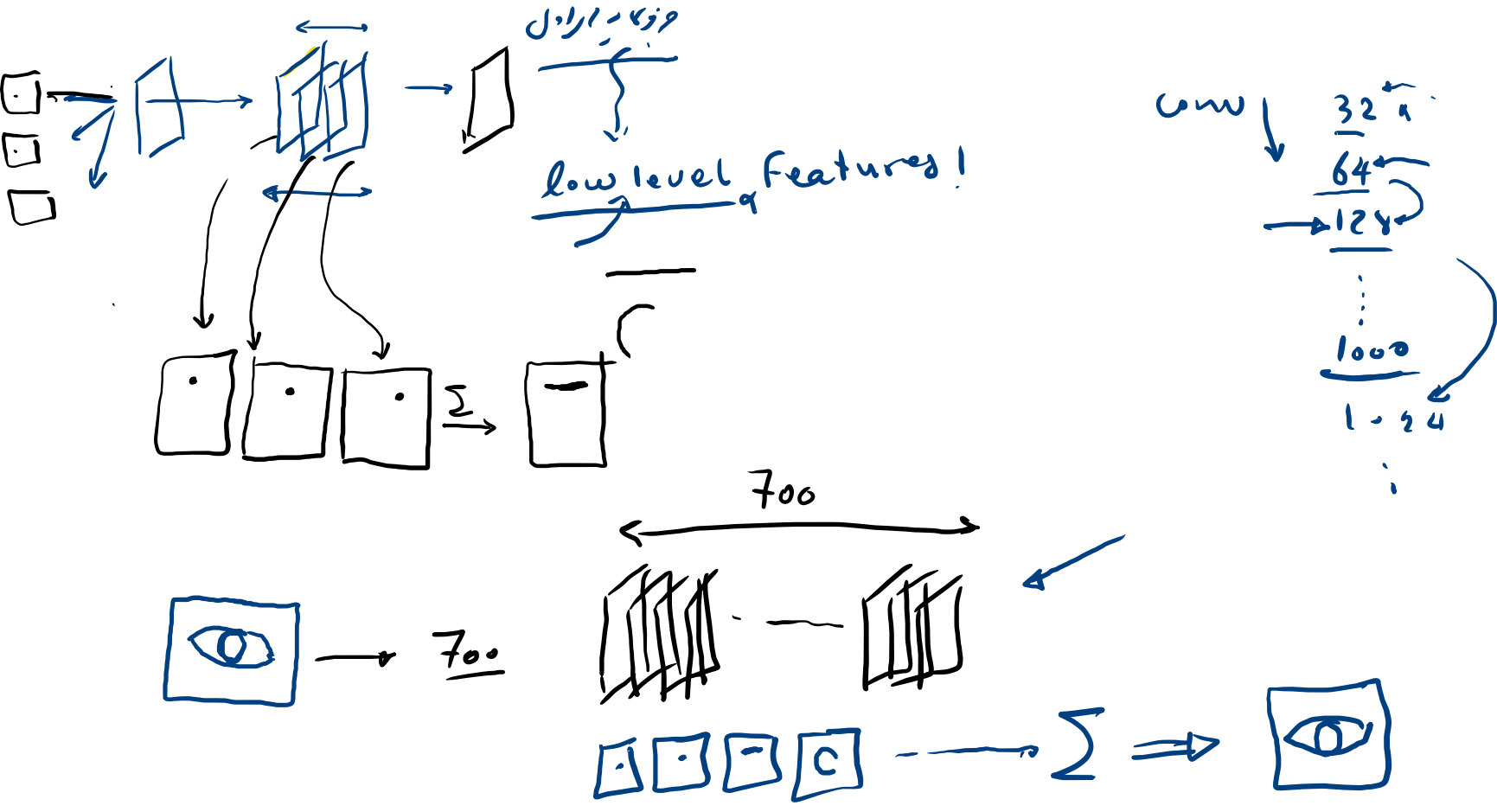
$$Sigmoid = \frac{1}{1 + e^{-net}}$$



$$\frac{\partial Sig(net)}{\partial net} \Rightarrow$$



max = ,25


low level Features!
700
700