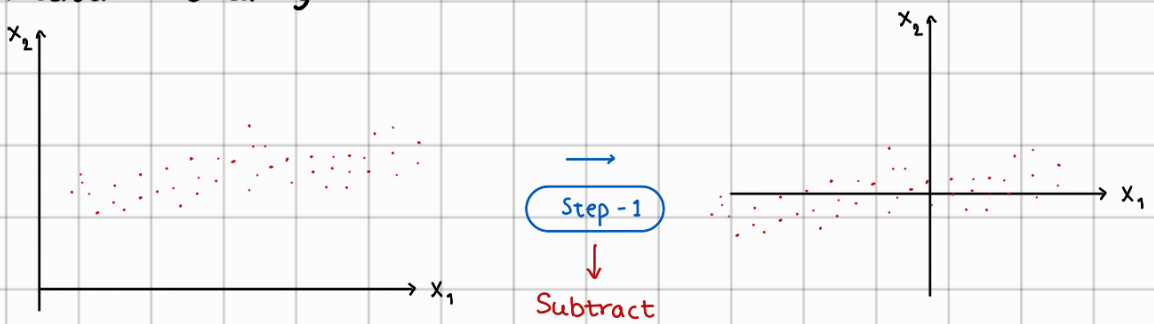
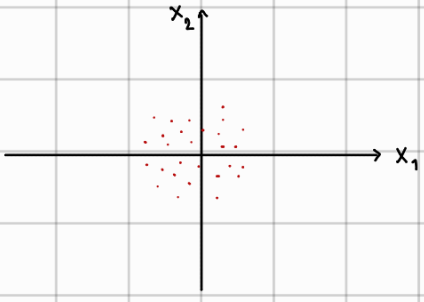
PHY654

Machine Learning in Particle Physics

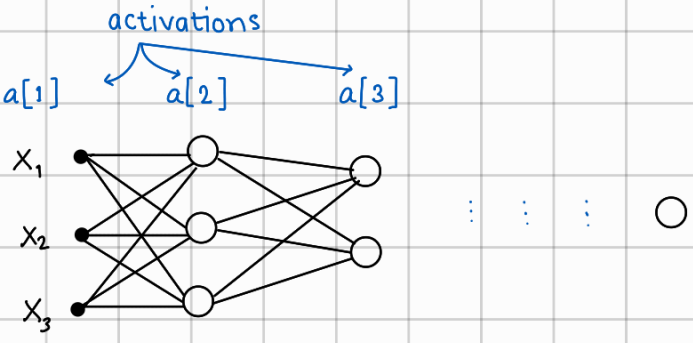
Saturday, 28 September, 2024

Feature Scaling:





Batch normalization:



$$Z^{[l]}$$

$$Z^{(l)}_{norm} = \frac{Z^{(l)} - \mu}{\sigma + \varepsilon}$$

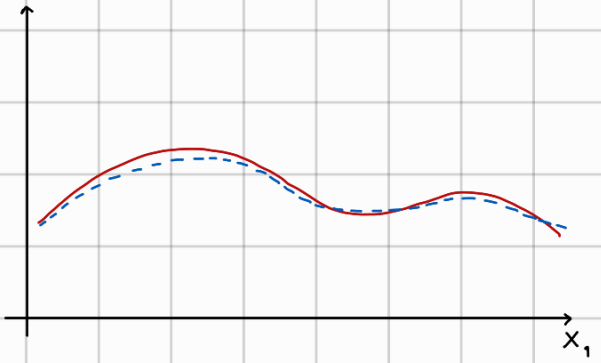
ε : a small number, just in case $\sigma = 0$

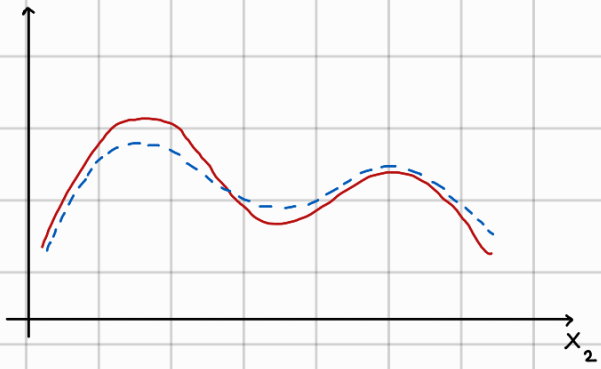
$$\tilde{Z}^{(l)} = \gamma Z^{(l)}_{norm} + \beta$$

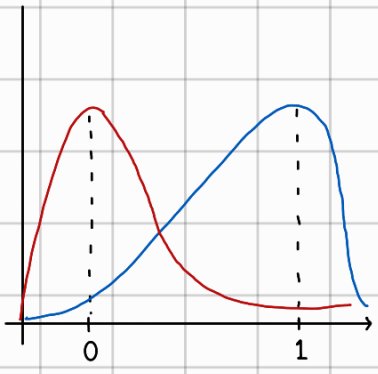
$\gamma = 1$, $\beta = 0$

model summary

conv-1	✓
pool	○
BN	✓





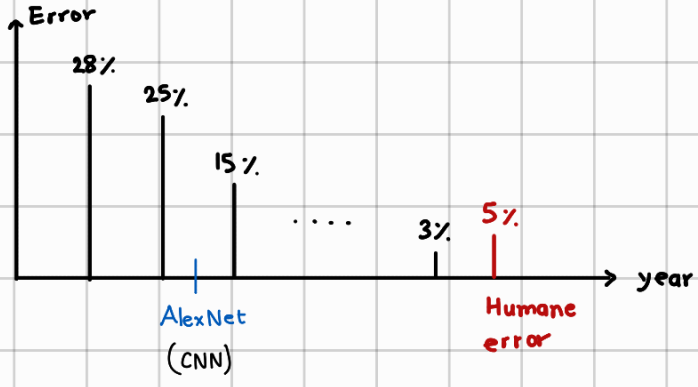


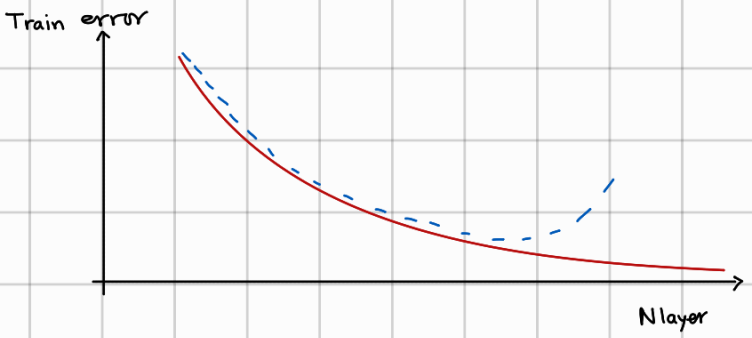
DNN Score

If there are 100 features, DNN looks in 100-dimensional plane for histograms.

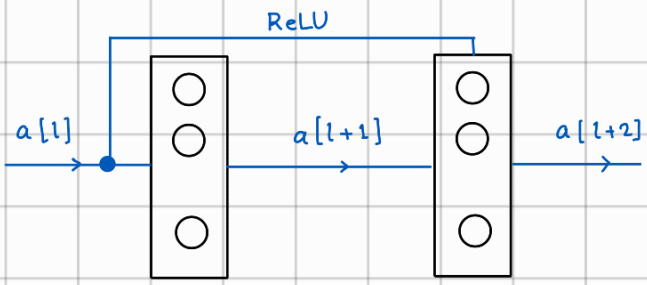
Resnet

152 layer CNN





Residual Block



$$z^{(l+1)} = w^{(l+1)} a^{(l)} + b^{(l+1)}$$

$$a^{(l+1)} = g(z^{(l+1)})$$

$$z^{(l+2)} = w^{(l+2)} a^{(l+1)} + b^{(l+2)}$$

Main Path

$$a^{(l+2)} = g(z^{(l+2)})$$

vs.

alternative path (skip connection)

$$a^{(l+2)} = g(z^{(l+2)} + a^{(l)})$$

$$a^{(l+2)} = g(z^{(l+2)} + K\, a^{(l)})$$

$a^{(l+2)}$: $n_1 \times 1$ $=$ $z^{(l+2)}$: $n_1 \times 1$ $+$ K: $n_1 \times n_2$ $\cdot$ $a^{(l)}$: $n_2 \times 1$

- Why ResNet works ?

$X \rightarrow$ DNN $\rightarrow a^{[l]}$

$X \rightarrow$ DNN $\rightarrow a^{[l]} \rightarrow a^{[l+2]}$

$$a^{[l+2]} = g(w^{[l+2]} a^{[l+1]} + b^{[l+2]} + a^{[l]})$$

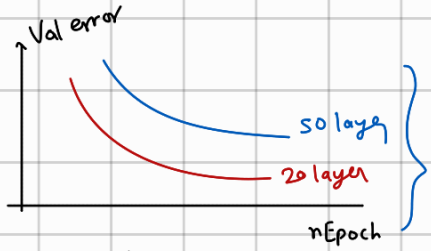
$b^{[l+2]}$: suppose it's 0 for simplicity

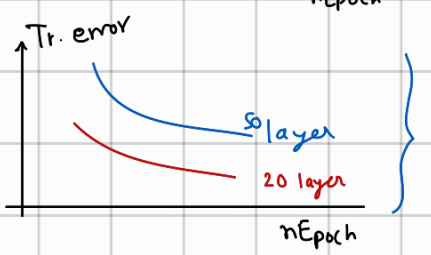
ReLU : $a^{[\cdot]} \geq 0$

(L2 Regularization)

$$a^{[l+2]} = a^{[l]}$$

(Vanishing gradient problem)





→ This is not due to overfitting

→ this is optimization issue.

X → input

$X_{Res} = X$

Conv-1 = conv2D( )(X)

Batch-1 = Batch 1 = B.N. ( ) (Conv-1)

act-1 = Activation('relu') (batch 1)

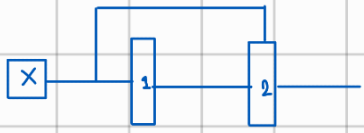
conv-2 = conv 2D ( ) (act 1)

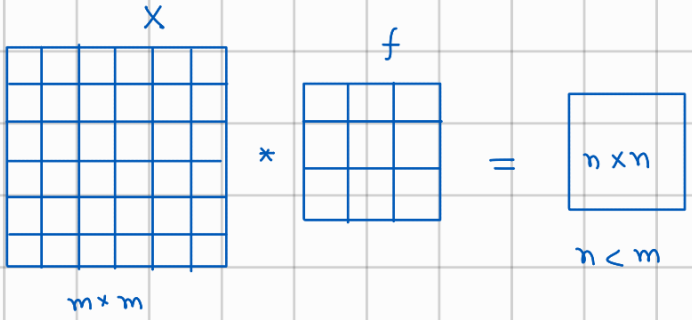
Batch-2 = BN ( ) (Conv-2)

ResNet = Add( ) (Batch_2, X_{Res})

Final = Act 'Relu' (ResNet)

~~Sequential~~





Conv2D Downsampling

Conv2D Transposed

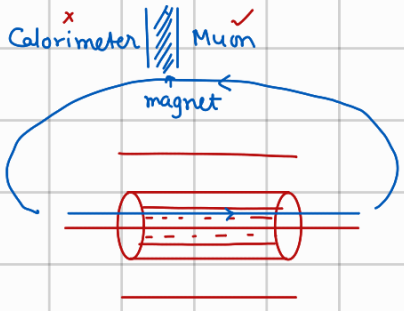
Theory level

p-p collision @ LHC
($\sqrt{s}$ = 13 TeV)

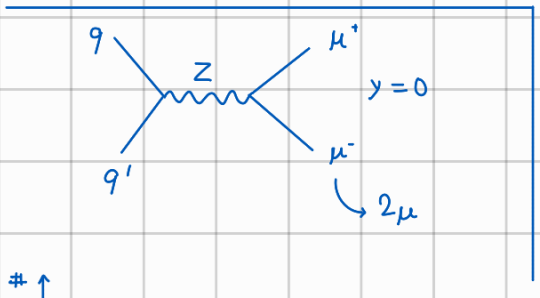
detector = CMS / ATLAS

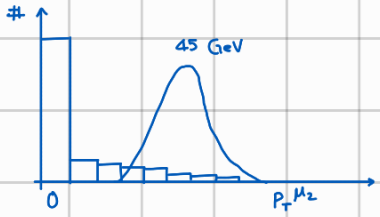
Detector Level

Tracker ✓ | Calorimeter ✗ | Muon ✓



μ + ME_T (missing transverse energy)





DNN

$X_1 = ME_T$ (float)

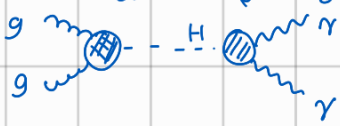
X_2 = nMuons (int)

$X_3 = p_T^{\mu_1}$

$X_4 = p_T^{\mu_2}$ (✓ or 0)

if found, if not exists

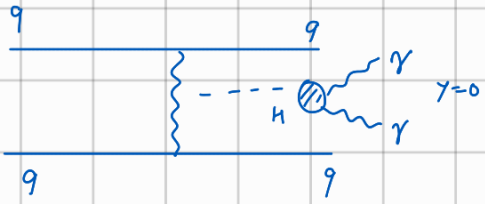
ggF (gluon-gluon fusion)



$y = 1$

DNN $x_1, x_2, \ldots$

VBF



$y = 0$