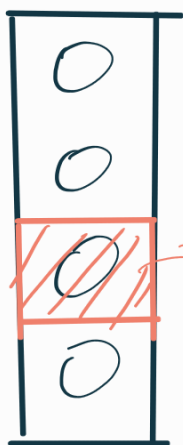


같은 layer의 feature map은 생성되면서 독립으로 나오게 되지만,
그 feature map들을 조합하여 새로운 feature map을 만드는
서로써 연관성 자체는 존재한다.

특, 독립적으로 생성은 되는데 다음 layer를 만들 때는 서로 연관성으로
재구성되므로 새로운 feature를 만들 때는, 종속적이다.

Feed-forward도 동일

Q, Batch - norm은 같은 layer의 다른 node의 상대적인
연관성을 고려하지 않는데 왜 성능이 좋을까?



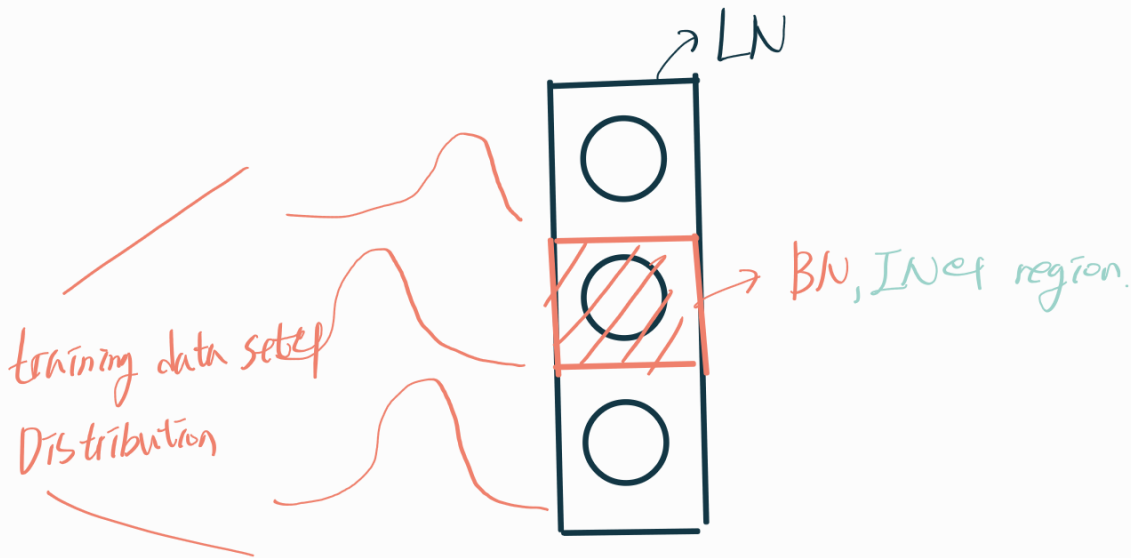
→ Instance - norm은 한 sample이 대해서 하기이, 상대적인 차이는
다양하게 고려 못하는 단점이 있지만 Batch - norm은 들어오는 input batch를
고려하니 학습이 진행됨에 따라 자연스럽게 상대적 차이를
(magnitude) 고려한다.



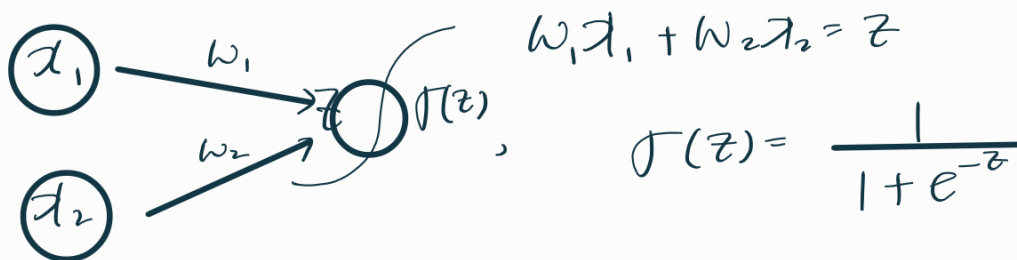
Instance-Norm의 γ, β 는 style에 대한 분포로 조절하는게지만

Batch-Norm은 training dataset의 Distribution을 조절하기 때문에
같은 layer에 다른 node가 서로 다른 차이를 고려할 수 있다.

Layer-Norm은 당연히 고려! (magnitude or 연관성)



Node가 독립이 아닌 이유 2.



$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \sigma} \frac{\partial \sigma}{\partial z} \frac{\partial z}{\partial w_1}$$

→ $\frac{e^{-z}}{(1 + e^{-z})^2}$ 인데, e^{-z} 이 x_2 가 들어가므로 영향을 준다.

x_1 이므로 x_2 에 영향 X

즉, 다른 feature 생성이 영향을 두므로 상대적으로 값이 중요하다.