

Introduction

Commands:

```
python app.py --data [xor | linear] --loss [mse | cross_entropy] --lr [float] --step [int] --unit 6  
6 [--use_activate] [--wandb]
```

if use wandb, you need to revise the code "wandb.init" to your own entity and project

--unit is array type, you can use --unit 1 2 3...n, number of hidden layers equals to length of unit

Basic Structure:

Model: Linear -> Relu -> Linear -> Relu -> Linear -> Sigmoid -> y

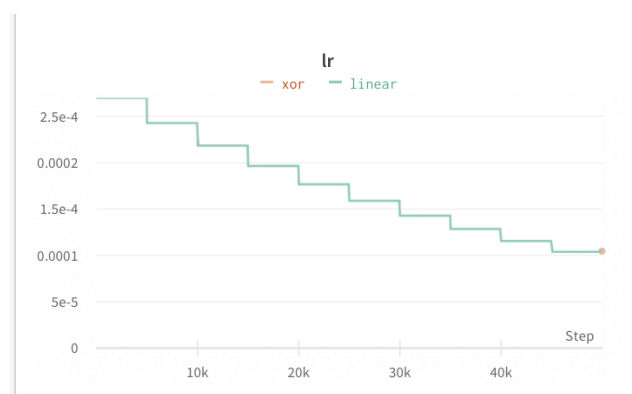
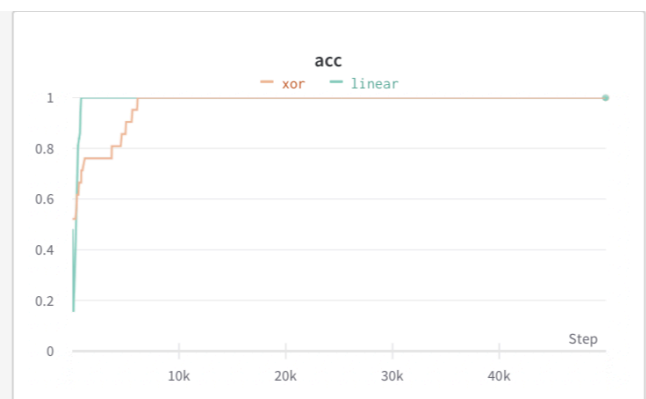
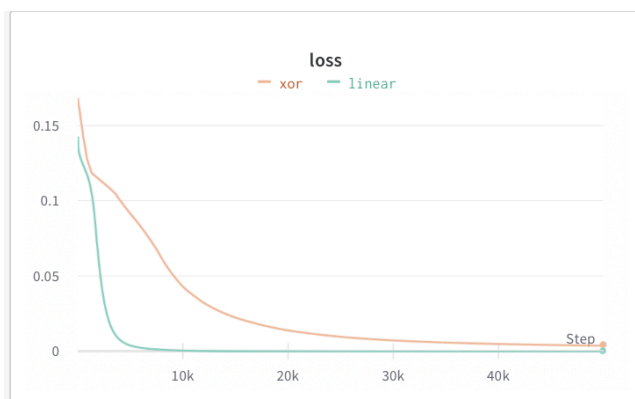
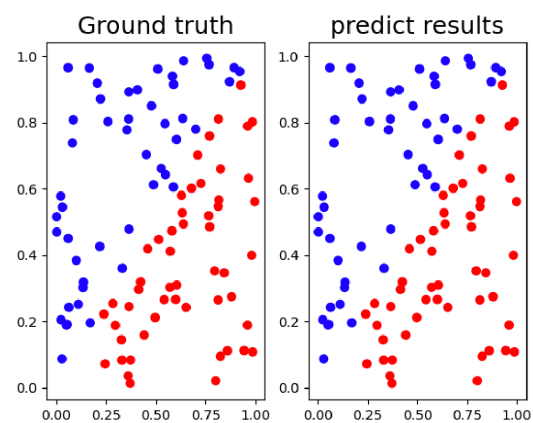
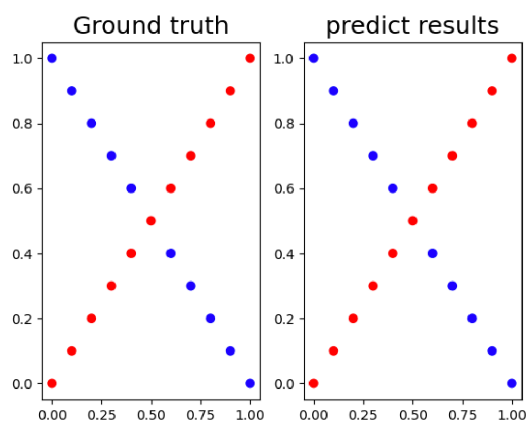
Loss: MSE

Channels: 1 -> 6 -> 6 -> 1

Step (epochs): 50000

Learning Rate: 0.0003 (Reduce 10% per 5000 steps)

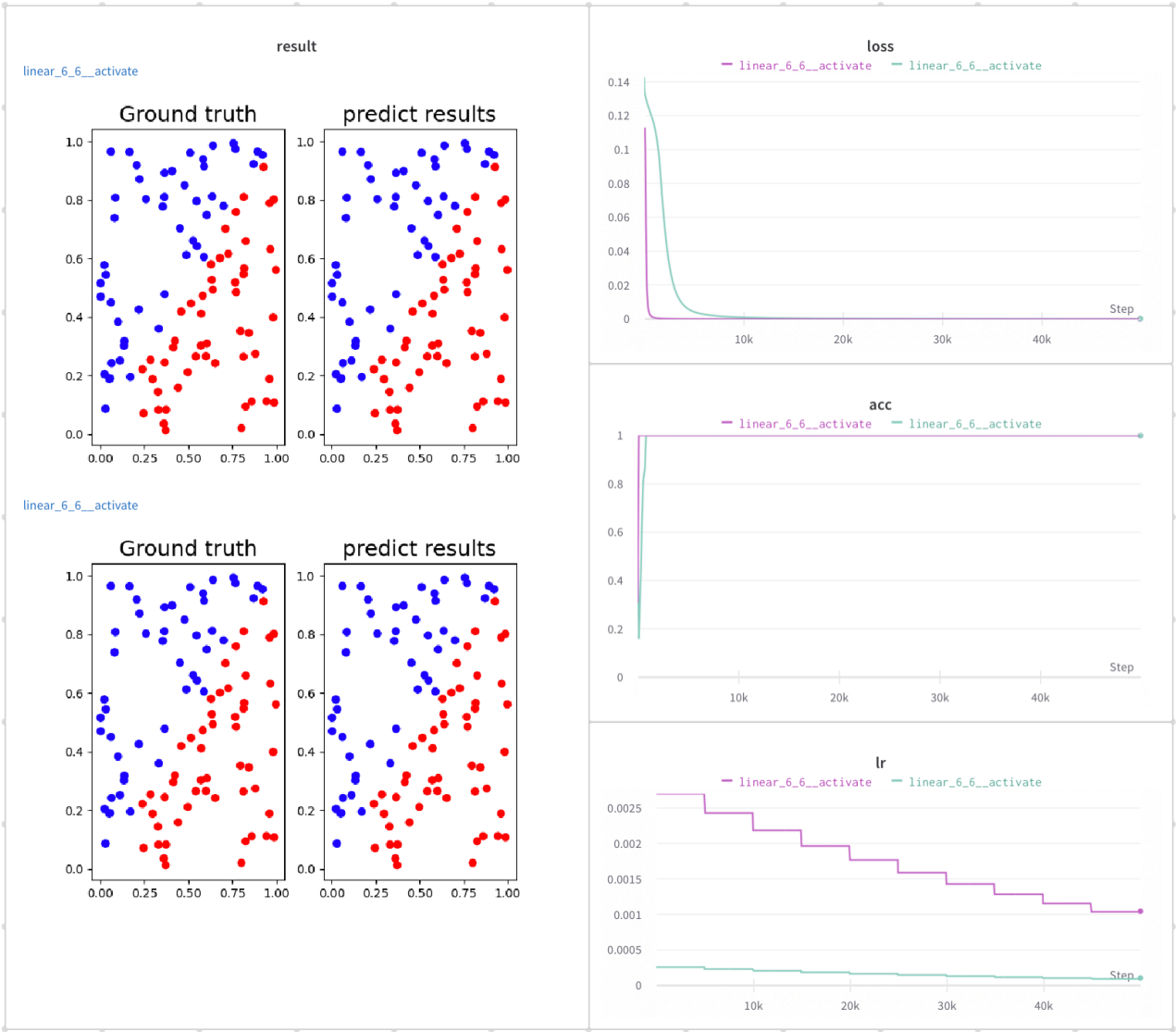
Result:



Experiment & Discussion

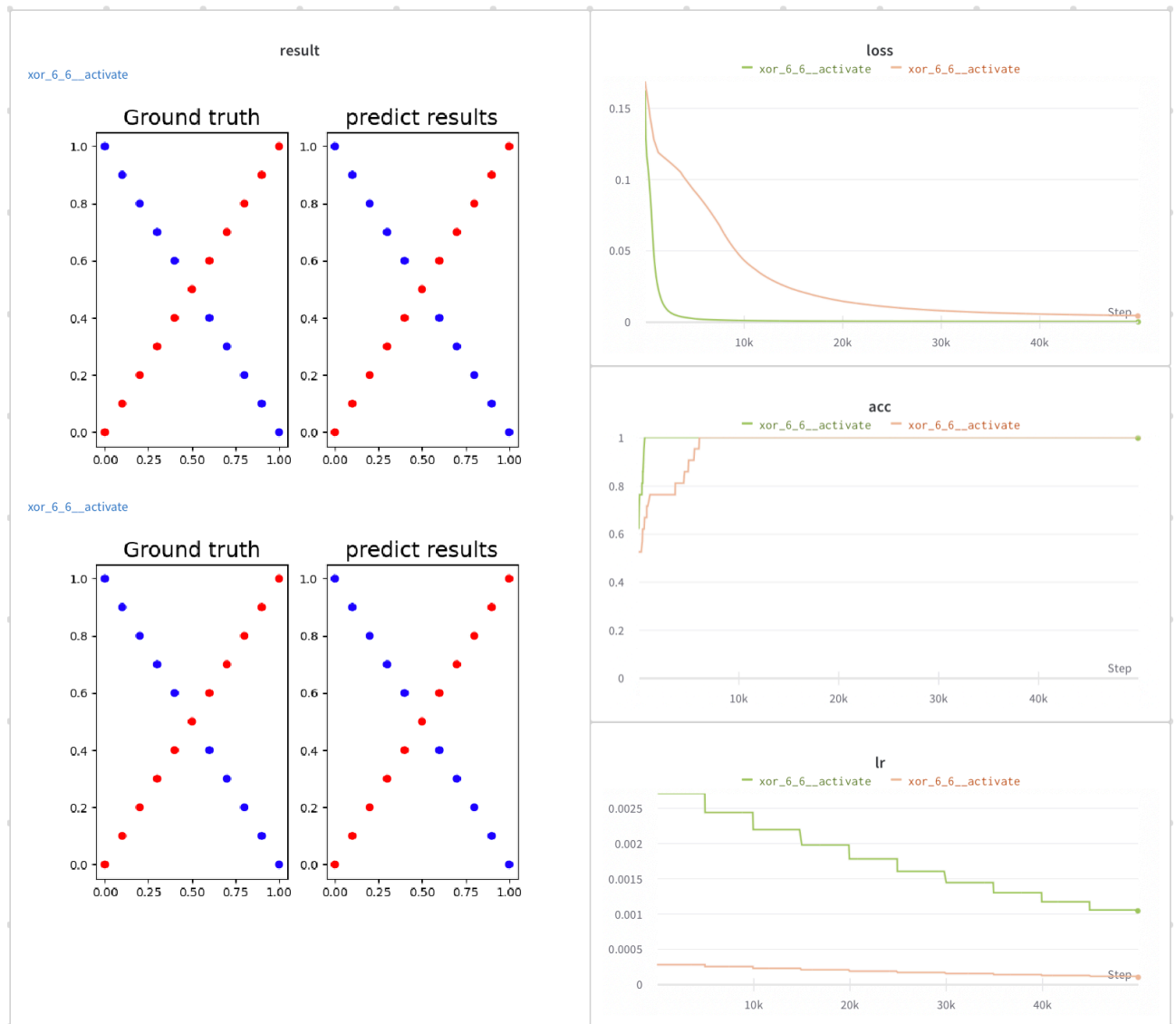
1. Different learning rate: 0.003 v.s 0.0003

Linear:



Name	Notes	data	loss	lr	unit	use_activat
●	Add notes...	linear	mse	0.003	[6,6]	true
●	control group	linear	mse	0.0003	[6,6]	true

XOR



●	Add notes...	xor	mse	0.003	[6,6]	true
●	control group	xor	mse	0.0003	[6,6]	true

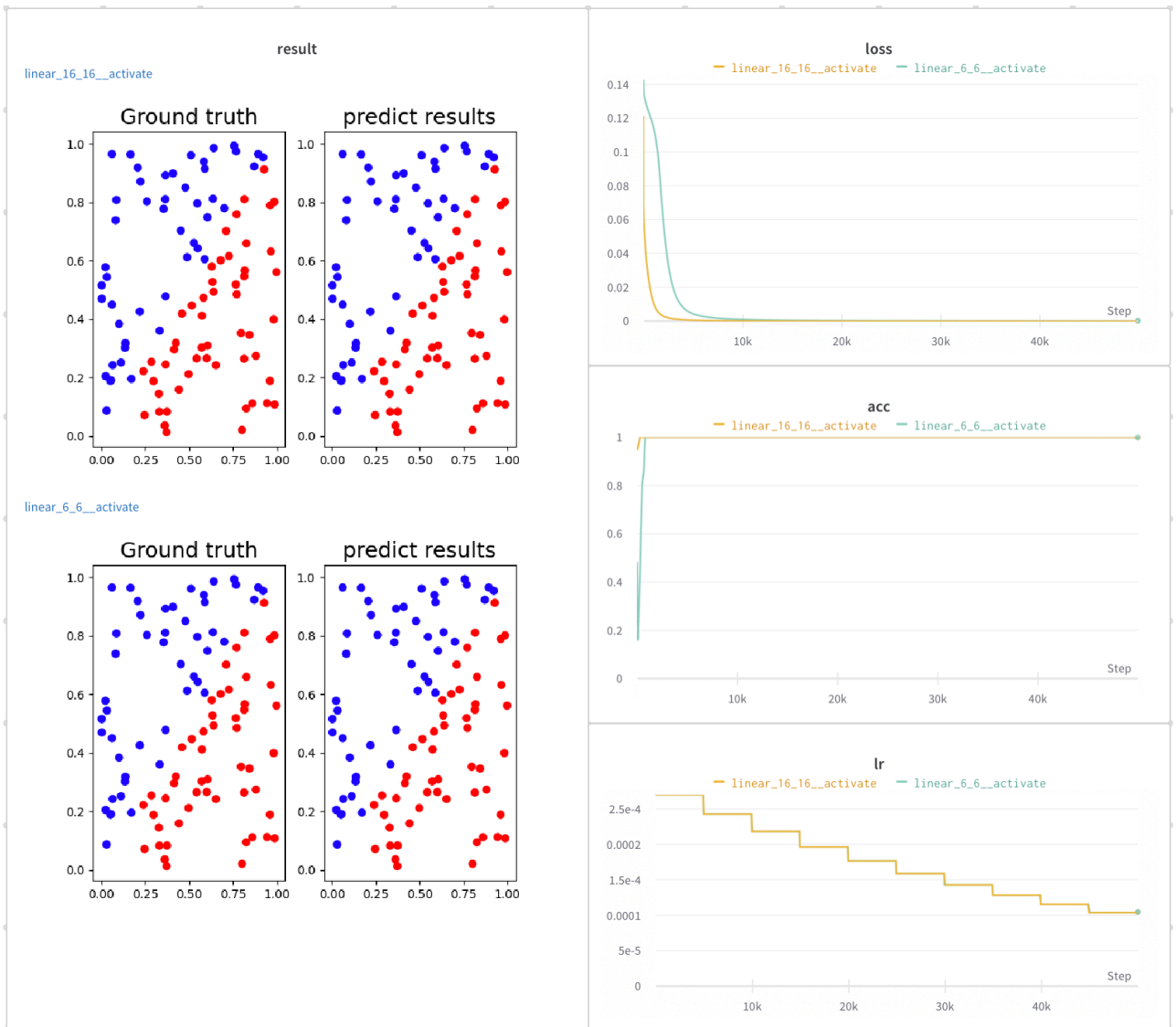
Learning Rate 與loss下降速度相關

如果 Learning Rate 越大 則 loss 下降速度越快

另外，如果Learning Rate太大 會導致 loss 上下起伏 無法收斂

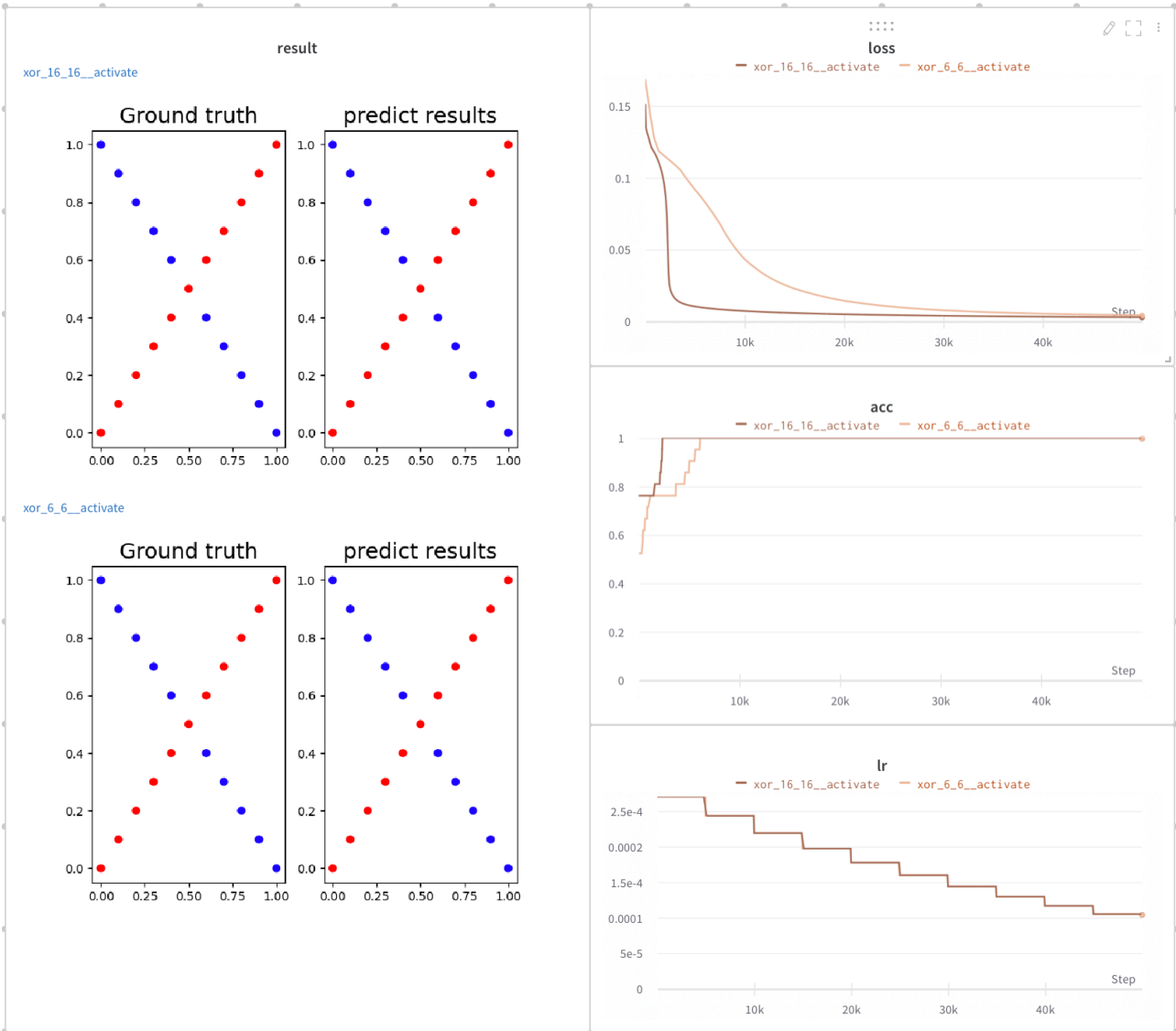
2. Different hidden unit size : [6,6] v.s [16,16]

Linear



Name	Notes	data	loss	lr	unit	use_activat
●	Add notes...	linear	mse	0.0003	[16,16]	true
●	control group	linear	mse	0.0003	[6,6]	true

XOR

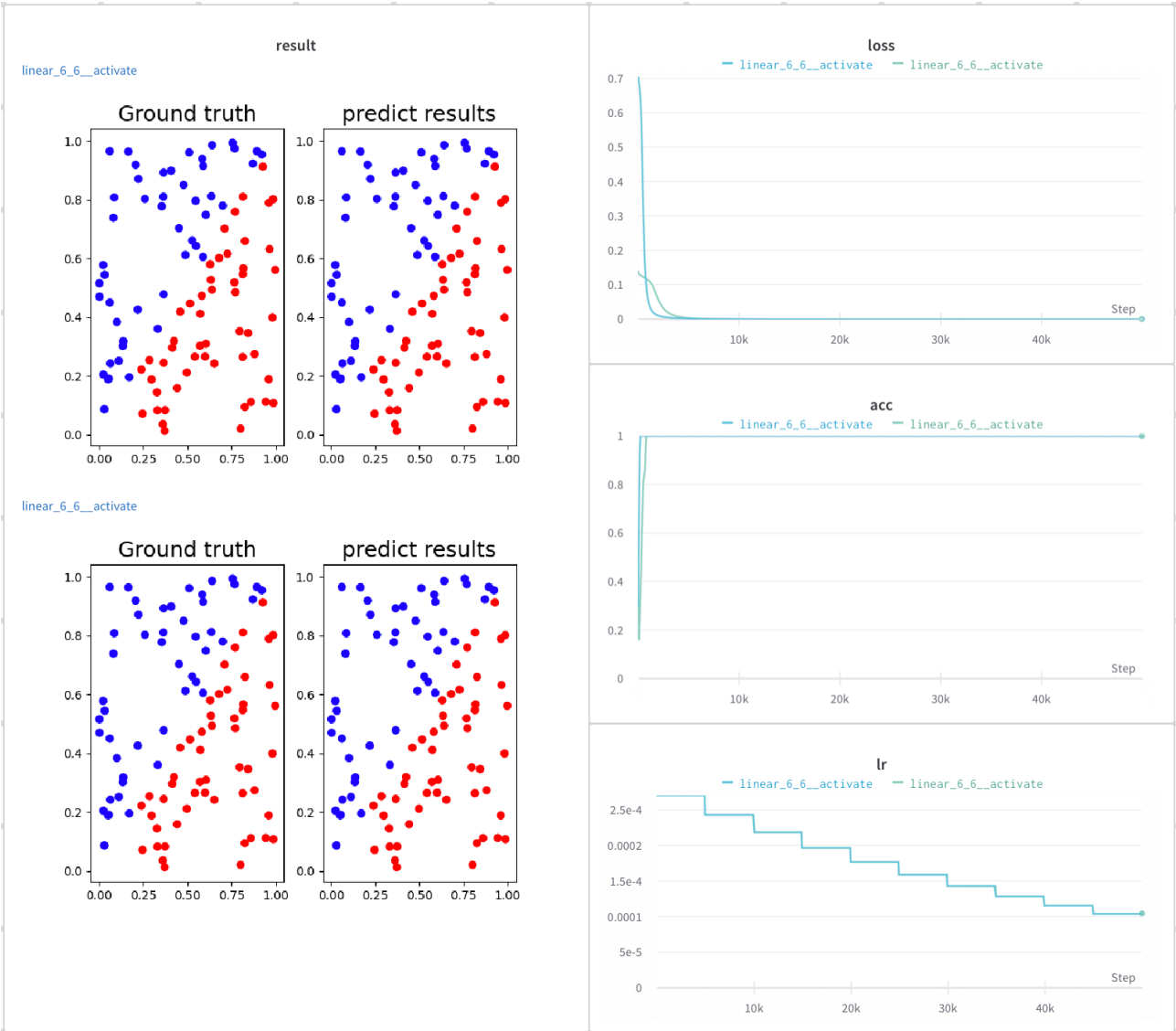


Name	Notes	data	loss	lr	unit :	use_activat
	Add notes...	xor	mse	0.0003	[16,16]	true
	control group	xor	mse	0.0003	[6,6]	true

在本次實驗中，每層 hidden unit 的數量越多，loss收斂的速度就會越快
意味著擁有比較強的model

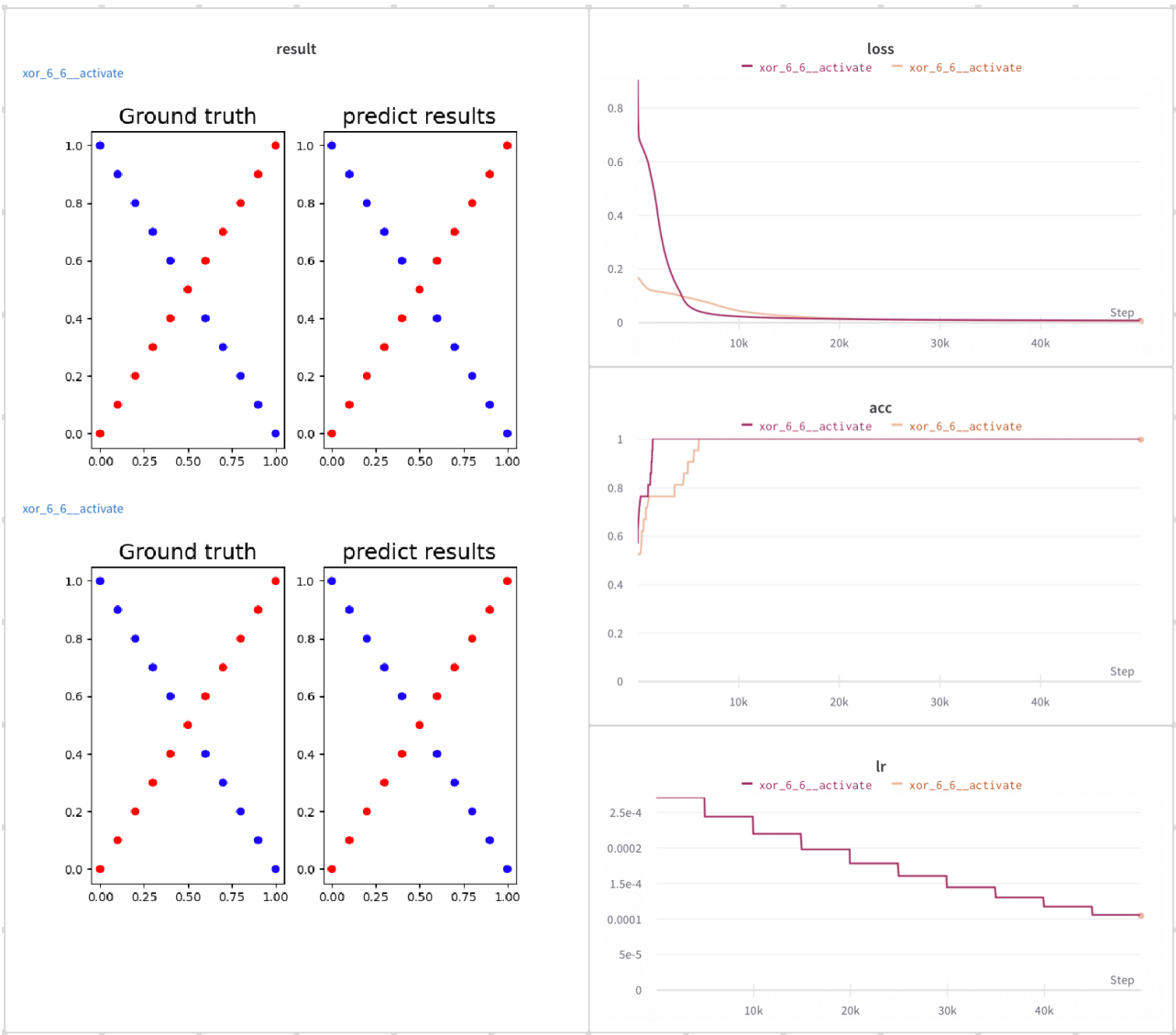
3. Different loss : MSE v.s Cross-entropy

Linear



Name	Notes	data	loss :	lr	unit	use_activat
●	Add notes...	linear	cross_entropy	0.0003	[6,6]	true
●	control group	linear	mse	0.0003	[6,6]	true

XOR

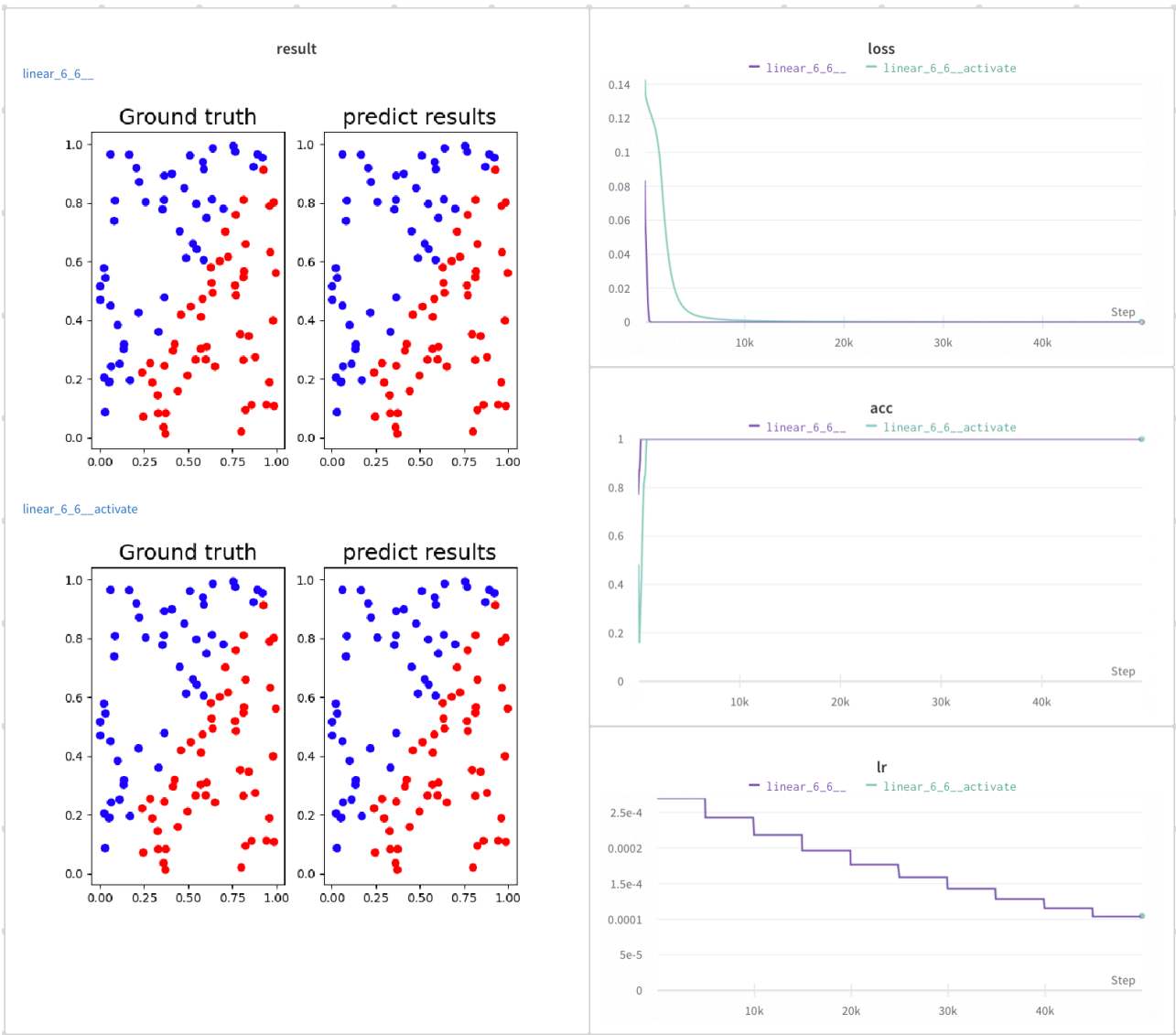


Name	Notes	data	loss :	lr	unit	use_activat
<div></div>	Add notes...	xor	cross_entropy	0.0003	[6,6]	true
<div></div>	control group	xor	mse	0.0003	[6,6]	true

在本次實驗中，Cross_entropy 收斂速度快於 MSE loss
在分類問題中，採用cross entropy是比較好的

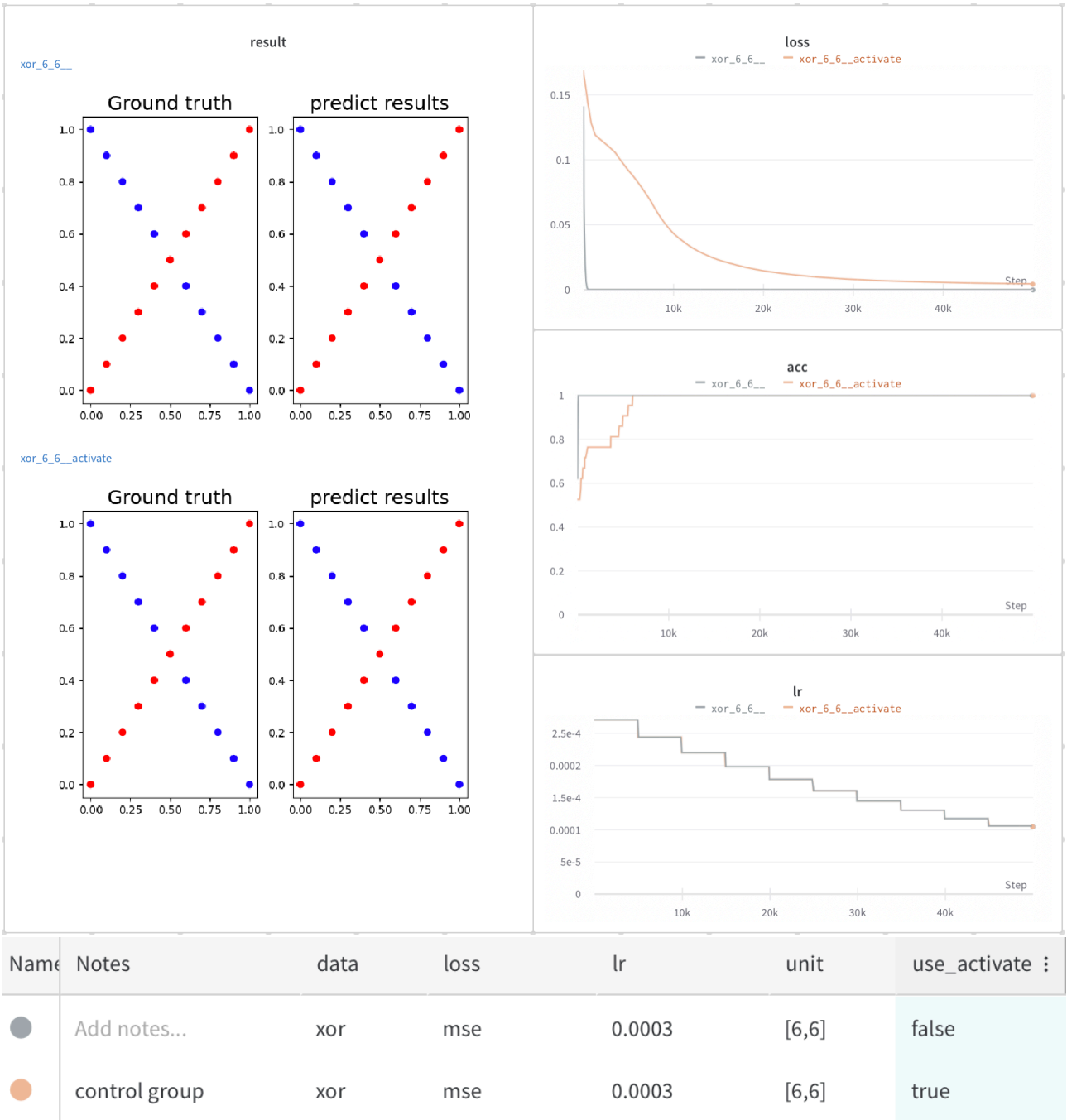
4. With/Without activate

Linear



> Name	Notes	data	loss	lr	unit	use_acti
> ●	Add notes...	linear	mse	0.0003	[6,6]	false
> ●	control group	linear	mse	0.0003	[6,6]	true

XOR



在本次實驗中，沒有activate function 的收斂速度較有activate function快很多
推測是MSE loss較適合 linear 的 output
此外 cross-entropy 無法使用 without activate function，會遇到 在0的時候 無法微分的情況