Notes on backprop for linear layer & vector tensor derivative linked to lecture 4 on syllabus

目前的进度：

Mini-batch SGD

Loop:

1. Smple a batch of data
2. Forward prop it through the graph(network) get loss
3. Backprop to calculate the gradients
4. Update the parameters using the gradients

下一步：

Training Neural Networks

1. **One time Setup:**

Activation functions

Pre-Processing,

Weight initial

Regularization,Gradient check

1. **Training dynamics:**

Babysitting the learning process 监控过程

Parameter updates

Hyperparameter optimization

1. **Evaluation**

Model ensembles

# Part 1

Activation function

Data pre-process

Weight initial

Batch normalization

Babysitting the training process

Hyper-parameter optimization

Activation Functions激活函数