



CSDN学院 IT实战派

# 第八章：录用之道-扎实的理论基础7

# 上一章回顾

- 评估指标
- 数据集的划分
- 交叉验证
- 贝叶斯极限与学习曲率
- 面试模拟-如何选择更优的模型

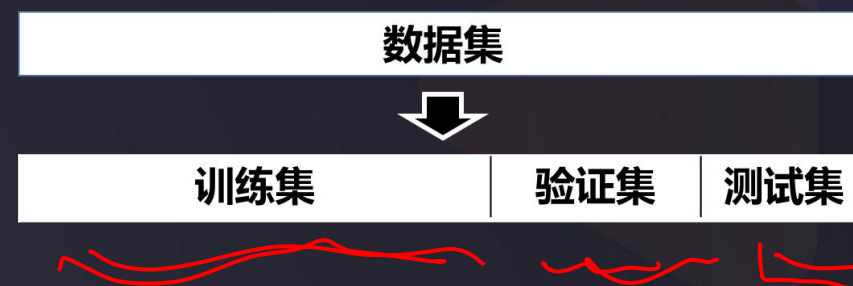
验证	训练	训练	训练
训练	验证	训练	训练
训练	训练	验证	训练
训练	训练	训练	验证

真实结果

正 负

预测结果	正	True Positive	False Positive
	负	False Negative	True Negative

1. Accuracy =  $(TP+TN)/(TP+FP+FN+TN)$
2. Precision =  $TP/(TP+FP)$
3. Recall =  $TP/(TP+FN)$



贝叶斯极限

人类识别准确度


训练准确度

验证准确度

测试准确度

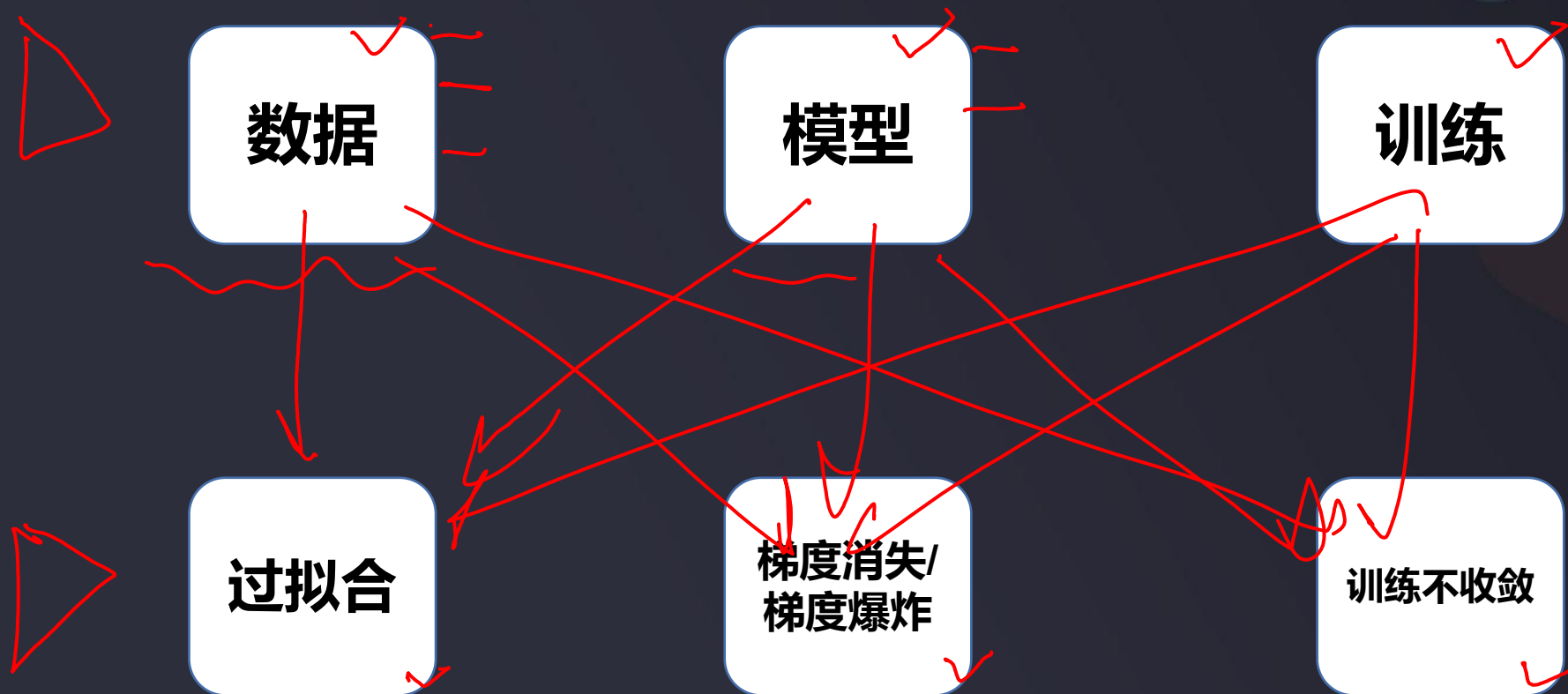
长时间真实应用环境下准确度

# I 深度学习模型优化

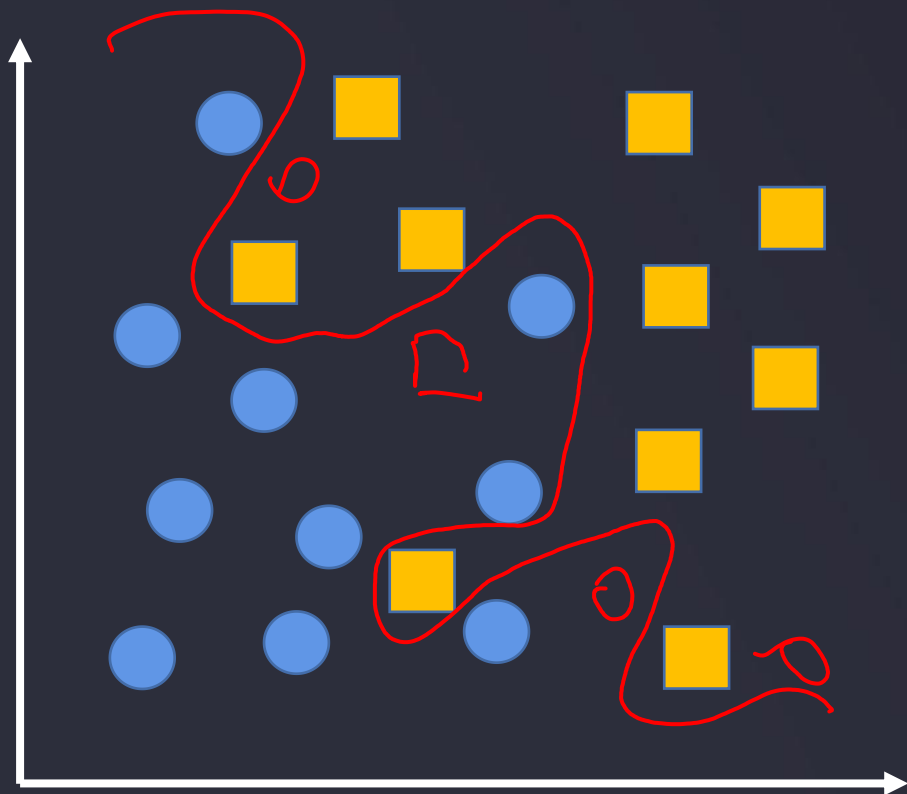


模型的机理	神经元与全连接神经网络	激活函数	权重与偏差
	卷积神经网络	循环神经网络	经典的模型架构
模型的训练	非监督学习	监督学习	强化学习
	损失函数	反向传播算法	泰勒展开
模型的评估	评估指标	贝叶斯极限	满足与优化指标
模型的优化	过拟合问题	训练的优化	梯度消失/梯度爆炸
	混合模型优化策略	其他优化策略	实战项目优化思路

# 深度学习模型优化



# 深度学习模型优化-过拟合问题



060  
060  
060



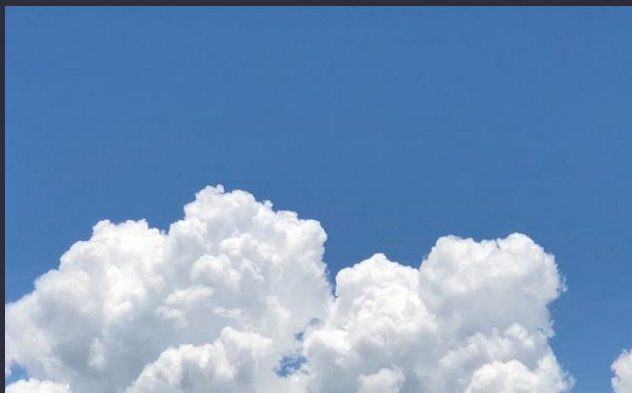
## 优化方法

- 1. 扩大数据集和代表性/数据增强 Data augmentation
- 2. 正则化 Regularization
- 3. 随机失活 Dropout

# 深度学习模型优化-扩大数据集和代表性



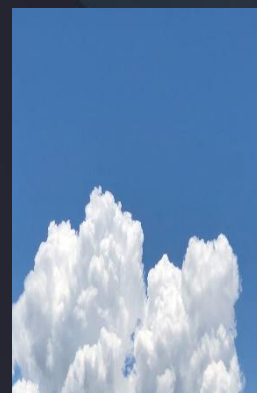
# 深度学习模型优化-数据增强 Data augmentation



旋转



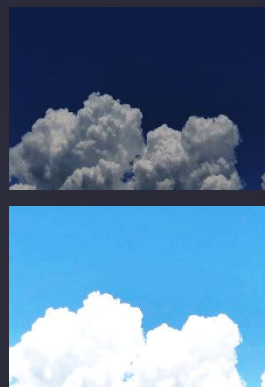
形变



镜像



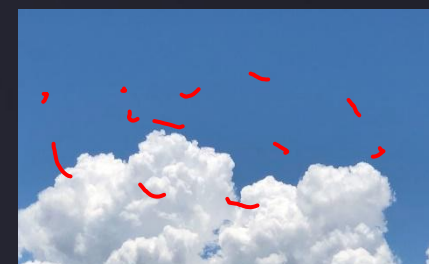
亮度



变色



噪音



# 深度学习模型优化-正则化Regularization

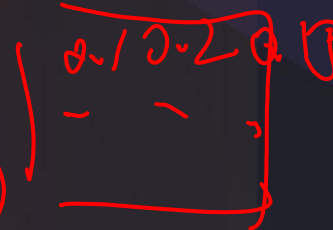
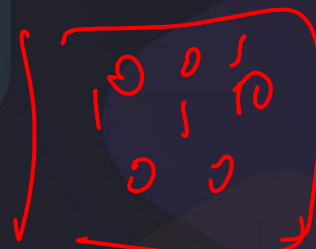
$f(x)$

$$L(Y, f(x)) \rightarrow L(Y, f(x), w)$$

$$L(Y, f(x)) = ||Y - f(x)||_2 + \alpha ||w||_2$$



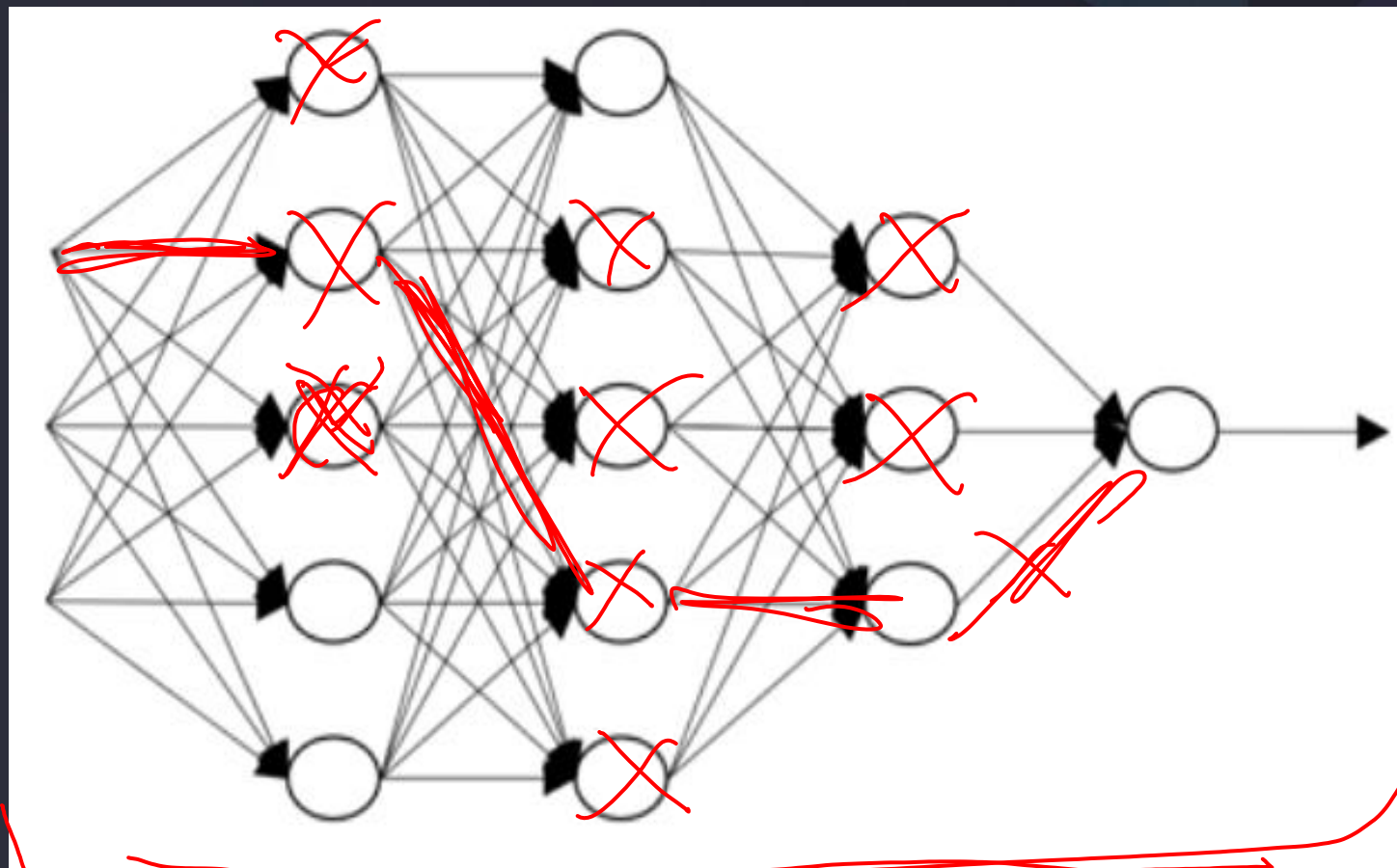
$f(\underline{w}x + \underline{b})$



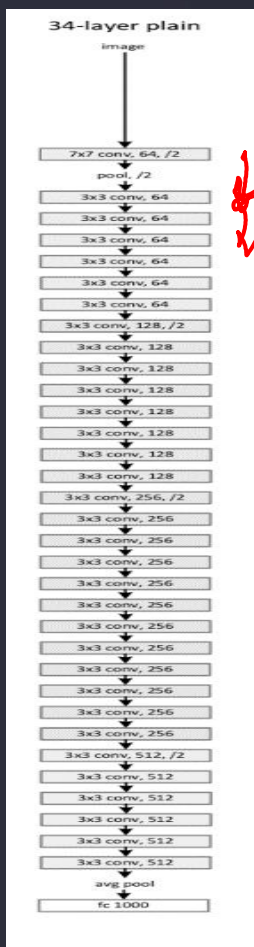
- L0范数: 向量中非零元素的个数, 记为 $||w||_0$
- L1范数: 绝对值之和, 记为 $||w||_1$
- L2范数: 平方和或者说模, 记为 $||w||_2$



# 深度学习模型优化-随机失活Dropout



# 深度学习模型优化-梯度消失与梯度爆炸问题



$$\frac{\partial C}{\partial b_1} = \frac{\partial C}{\partial y_4} \frac{\partial y_4}{\partial z_4} \frac{\partial z_4}{\partial x_4} \frac{\partial x_4}{\partial z_3} \frac{\partial z_3}{\partial x_3} \frac{\partial x_3}{\partial z_2} \frac{\partial z_2}{\partial x_2} \frac{\partial x_2}{\partial z_1} \frac{\partial z_1}{\partial b_1}$$

Handwritten notes:

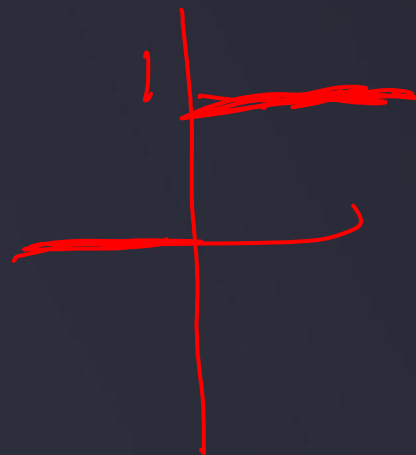
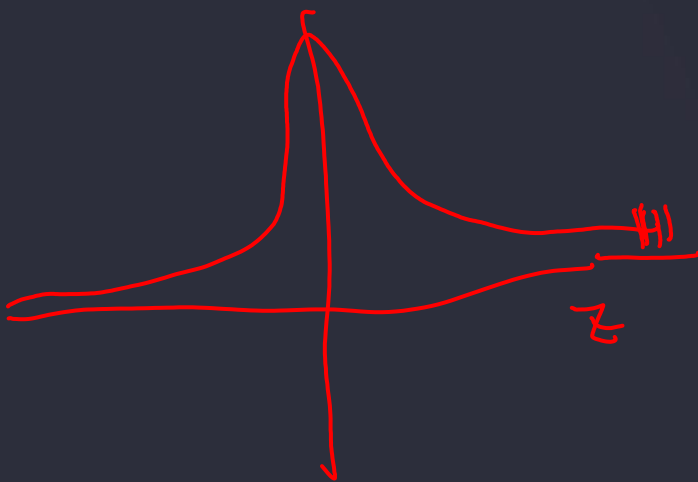
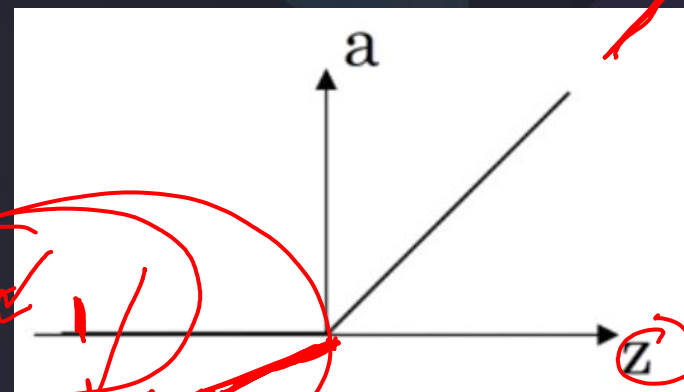
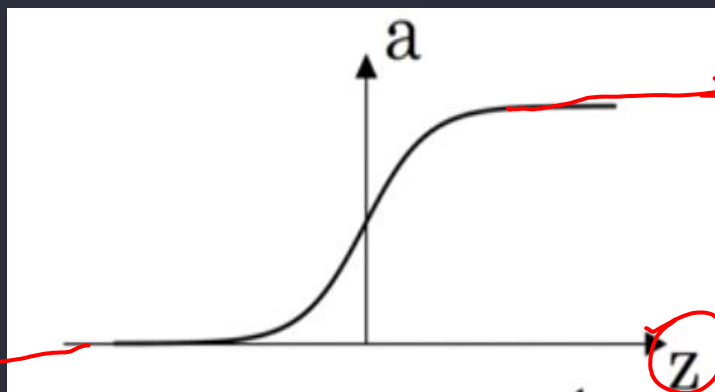
$$2^{10} = 1024$$

$$\frac{\partial J^{(0)}}{\partial w} = \frac{1}{1024}$$

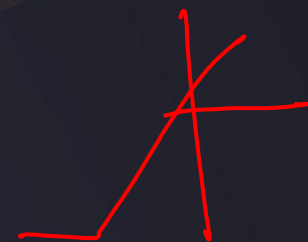
## 优化方法

- 1. 残差神经网络连接方式
2. 非饱和的激活函数 (如 ReLU)
3. 梯度截断 (Gradient Clipping)
4. 好的参数初始化方式
5. 批量规范化 (Batch Normalization)
- 6. 更快的优化器
- 7. LSTM

# 深度学习模型优化-非饱和的激活函数 (ReLU) 与梯度截断 (Gradient Clipping)



$$\begin{aligned} & > \theta & \theta \\ & < -\theta & -\theta \end{aligned}$$

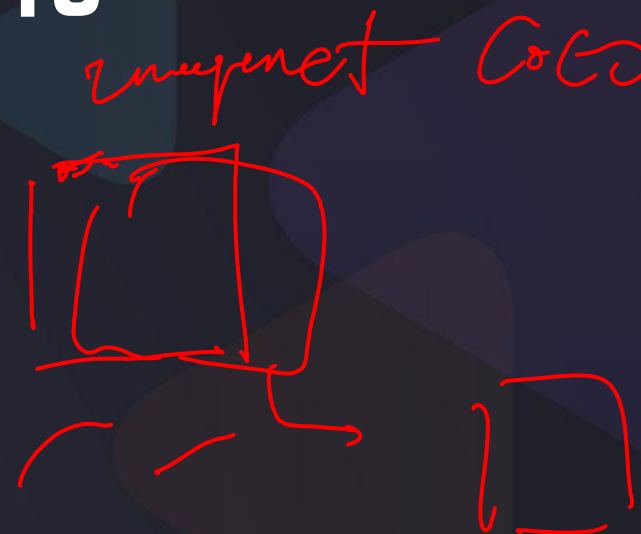


# 深度学习模型优化-网络参数的初始化

初始化的方法:

1. ~~全零/全一~~
2. ~~随机初始化~~
3. ~~Xavier初始化~~
4. He初始化
5. Pre-train初始化 (迁移学习)

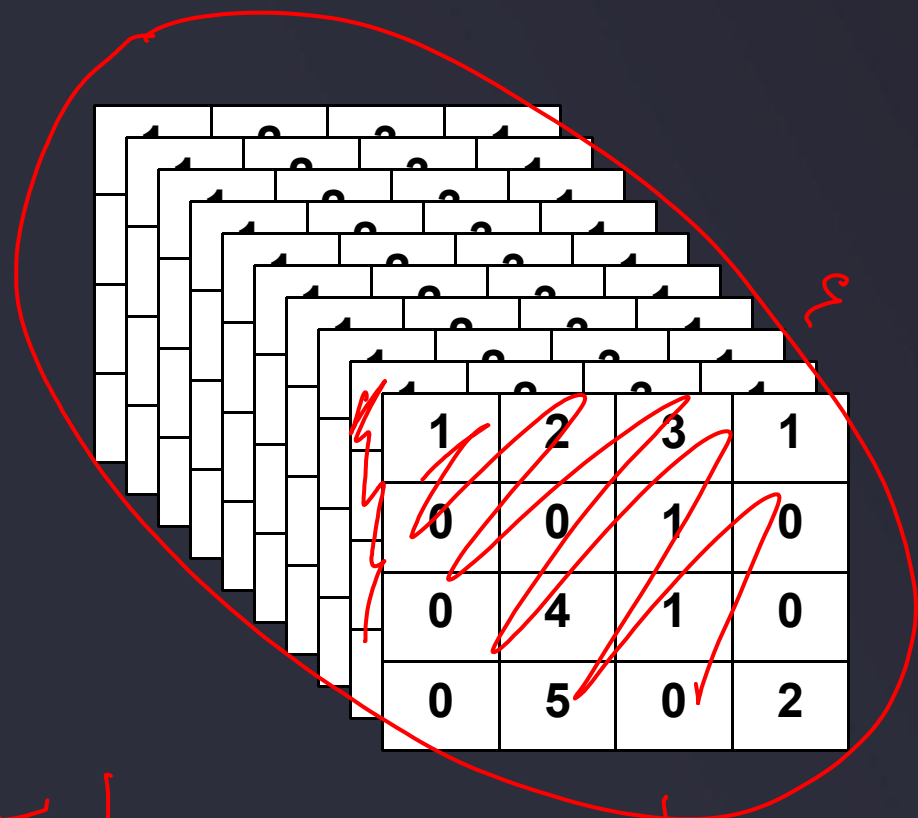
$$W \sim U\left[-\frac{\sqrt{6} \times 2}{\sqrt{n_j + n_{j+1}}}, \frac{\sqrt{6} \sqrt{2}}{\sqrt{n_j + n_{j+1}}}\right]$$



ReLU

He

# 深度学习模型优化-批量规范化 (Batch Normalization)



1. 对防止gradient vanish
2. 解决了Internal Covariate Shift的问题,可以提高学习速率
3. 减少了对于好的权重初始化的依赖
4. 有助于解决overfitting

# 深度学习模型优化-训练未收敛



1. 调整学习速率
2. 批训练
3. 优化器的使用

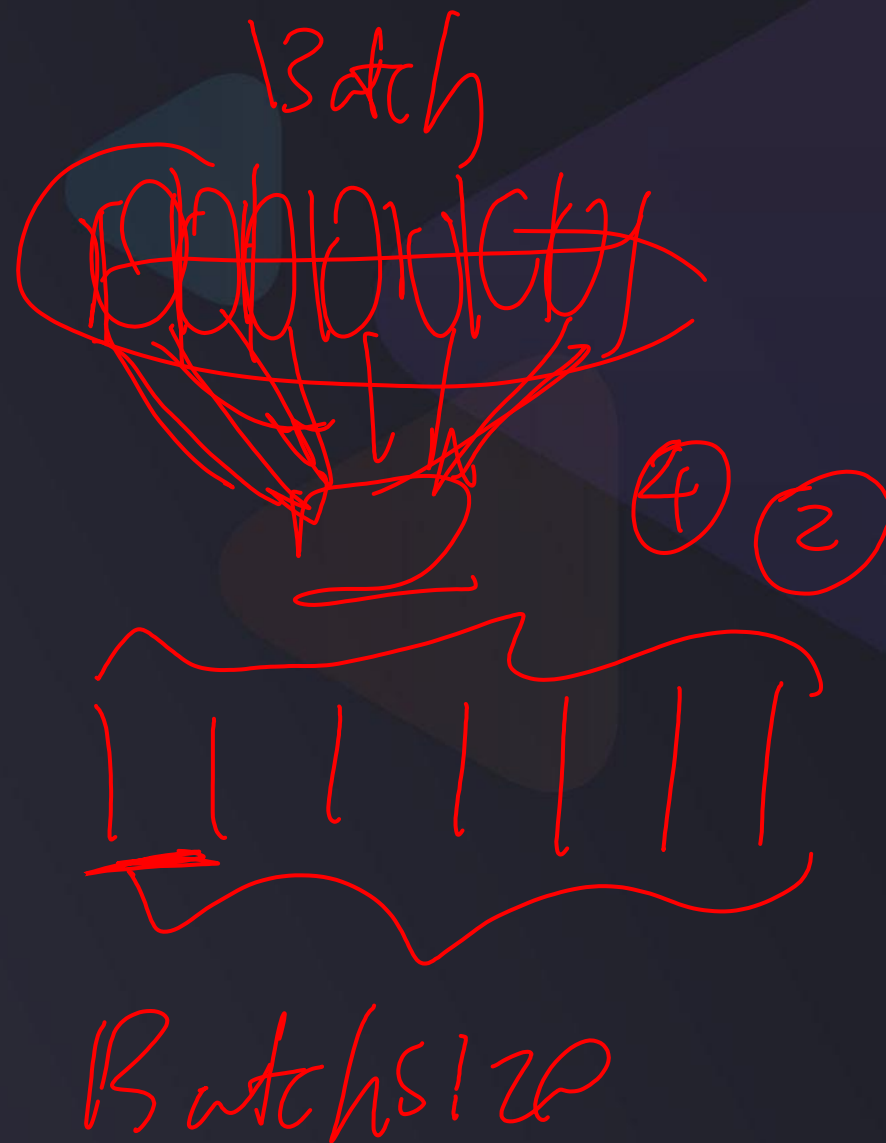
# 深度学习模型优化-学习速率优化

- 1.调整学习速率
- 2.学习速率递减



# 深度学习模型优化-批训练

1. 提高训练速度
2. 对训练过程引入随机性





# 深度学习模型优化-动量梯度下降法(gradient descent with momentum)优化器

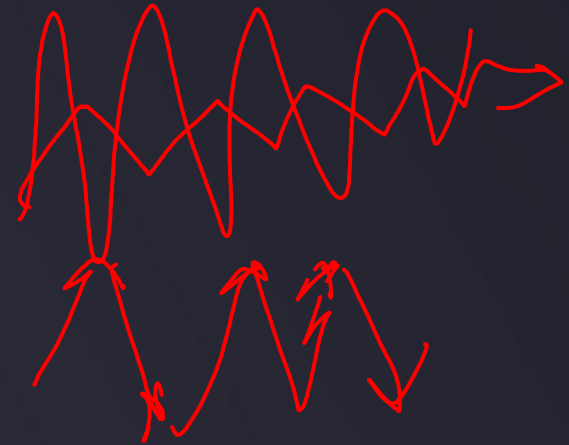
$$v_w = \beta v_w + (1 - \beta) dW$$

$$v_b = \beta v_b + (1 - \beta) db$$

$$W = W - \alpha v_w \quad b = b - \alpha v_b$$

$$W - \alpha dW$$

$$W - \alpha v_w$$



# 深度学习模型优化-RMSProp优化器

$$\begin{aligned} S_w &= \beta S_w + (1 - \beta) dW^2 \\ S_b &= \beta S_b + (1 - \beta) db^2 \\ W &= W - \alpha \frac{dW}{\sqrt{S_w}} \\ b &= b - \alpha \frac{db}{\sqrt{S_b}} \end{aligned}$$

# 深度学习模型优化-自适应矩估计Adam优化器

0-1

$$v_w = \beta_1 v_w + (1 - \beta_1) dW$$

$$v_b = \beta_1 v_b + (1 - \beta_1) db$$

$$S_w = \beta_2 S_w + (1 - \beta_2) dW^2$$

$$S_b = \beta_2 S_b + (1 - \beta_2) db^2$$

$$v_w = v_w / (1 - \beta_1^t) \quad v_b = v_b / (1 - \beta_1^t)$$

$$s_w = s_w / (1 - \beta_2^t) \quad s_b = s_b / (1 - \beta_2^t)$$

$$W = W - \alpha \frac{V_w}{\sqrt{(S_w) + \epsilon}} \quad b = b - \alpha \frac{V_b}{\sqrt{(S_b) + \epsilon}}$$

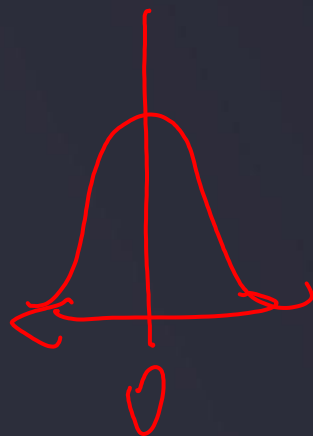


Kingma  
Le: Ba

$\beta_1 = 0.9$   
 $\beta_2 = 0.999$   
 $\epsilon = 10^{-8} / 9$

# 深度学习模型优化-其他优化策略

- 数据清洗/数据归一化 ←
- 非对称数据优化 ←
- 超参数优化 ←



Eng'

Na N

## 面试模拟-为何要对数据做归一化处理

# 面试模拟-为何要对数据做归一化处理

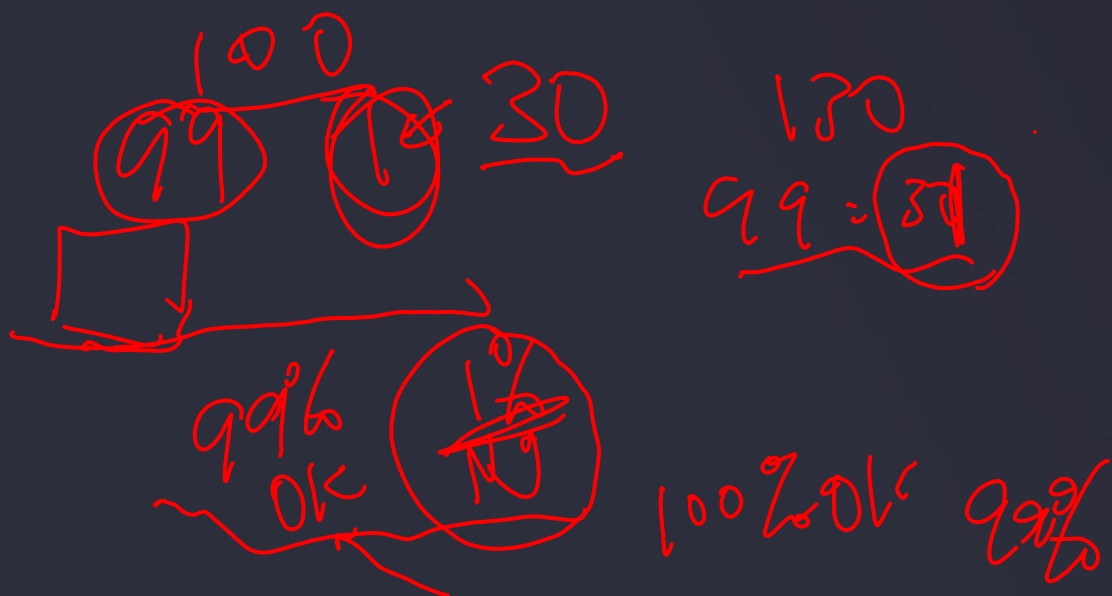
1. 提高训练速度, 帮助训练收敛, 避免梯度消失/爆炸
2. 帮助模型专注数据真正的关联性, 提高模型最终表现

$$\frac{x - \bar{x}}{\delta}$$

$$\frac{x - \bar{x}}{\text{Max}(x) - \text{Min}(x)}$$

# 面试模拟-非对称数据训练和优化

- ▷ 数据增强扩大比例较低的样本的数量
- 修改损失函数赋予比例较低样本更高的权重



$$L(S_{\text{data}}, WS_{\text{data}})$$

# 面试模拟-超参数的优化

$W, b$

$\beta_1, \beta_2$

loss

sgd

sgd/Adam

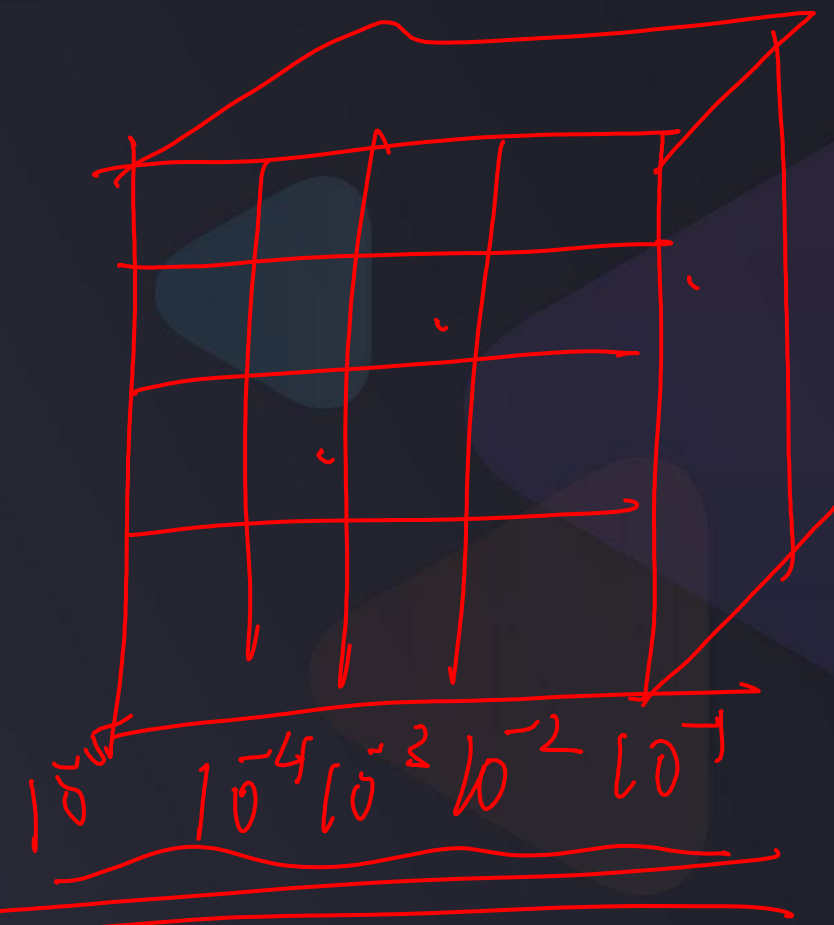
...

$2^n$   $10^n$

$M$

进化算法

Genetic Algorithm



→ 1. 快速实现  
2. 迭代试错



# 面试模拟-超参数的优化

1. 快速把模型搭建起来，初步跑通获得识别结果
2. 对识别结果进行分析，然后反复迭代优化（结合行业经验），直到获得一个满意的结果
3. 通过数量级搜索或者进化算法对第二步进行提速

理解

# 本章回顾

过拟合



数据



1. 扩大数据集
2. 数据增强
3. 数据清洗
4. 归一化

梯度消失/  
梯度爆炸



模型



1. 正则化/修改损失函数
2. 激活函数选取 (如 ReLU)
3. 特定结构(ResNet/LSTM)
4. 超参数的优化

训练不收敛



训练



1. 随机失活
2. 梯度截断 (Gradient Clipping)
3. 好的参数初始化方式
4. 批训练
5. 批量规范化 (Batch Normalization)
6. 优化器

# 如何接手一个实战项目

1. 定义任务和目标-选取好的模型以及结构超参数
2. 对模型进行训练
3. 评估模型的表现-发现问题（过拟合/梯度消失）
4. 根据模型表现进行优化

## 课程相关资料





欢迎大家扫码或者添加微信好友ai\_flare（学习小助手），加入学习群，老师会在群里帮大家解答学习、职业发展与求职问题（名额有限、人满即止）