

# 法律声明

本课件包括演示文稿、示例、代码、题库、视频和声音等内容，深度之眼和讲师拥有完全知识产权；只限于善意学习者在本课程使用，不得在课程范围外向任何第三方散播。任何其他人或者机构不得盗版、复制、仿造其中的创意和内容，我们保留一切通过法律手段追究违反者的权利。

## 课程详情请咨询

- 微信公众号：深度之眼
- 客服微信号：deepshare0920



公众号



微信

关注公众号深度之眼，后台回复论文，获取60篇AI必读经典前沿论文





deepshare.net

深度之眼

# Batch Normalization

导师：余老师

---

关注公众号深度之眼，后台回复论文，获取60篇AI必读经典前沿论文



# 目录

1/ Batch Normalization 概念

2/ PyTorch的Batch Normalization 1d/2d/3d 实现



# Batch Normalization

Batch Normalization

---

Batch Normalization: **批标准化**

**批**: 一批数据, 通常为mini-batch

**标准化**: 0均值, 1方差

优点:

1. 可以**用**更大学习率, 加速模型收敛
2. 可以**不用**精心设计权值初始化
3. 可以**不用**dropout或较小的dropout
4. 可以**不用**L2或者较小的weight decay
5. 可以**不用**LRN(local response normalization)

《Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift》

关注公众号深度之眼, 后台回复论文, 获取60篇AI必读经典前沿论文

# Batch Normalization

Batch Normalization

## 计算方式

affine transform

增强Capacity

**Input:** Values of  $x$  over a mini-batch:  $\mathcal{B} = \{x_{1...m}\}$ ;

Parameters to be learned:  $\gamma, \beta$

**Output:**  $\{y_i = \text{BN}_{\gamma, \beta}(x_i)\}$

$$\mu_{\mathcal{B}} \leftarrow \frac{1}{m} \sum_{i=1}^m x_i \quad // \text{ mini-batch mean}$$

$$\sigma_{\mathcal{B}}^2 \leftarrow \frac{1}{m} \sum_{i=1}^m (x_i - \mu_{\mathcal{B}})^2 \quad // \text{ mini-batch variance}$$

$$\hat{x}_i \leftarrow \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}} \quad // \text{ normalize}$$

$$y_i \leftarrow \gamma \hat{x}_i + \beta \equiv \text{BN}_{\gamma, \beta}(x_i) \quad // \text{ scale and shift}$$

**Algorithm 1:** Batch Normalizing Transform, applied to activation  $x$  over a mini-batch.

# Batch Normalization

Batch Normalization

## Internal Covariate Shift (ICS)

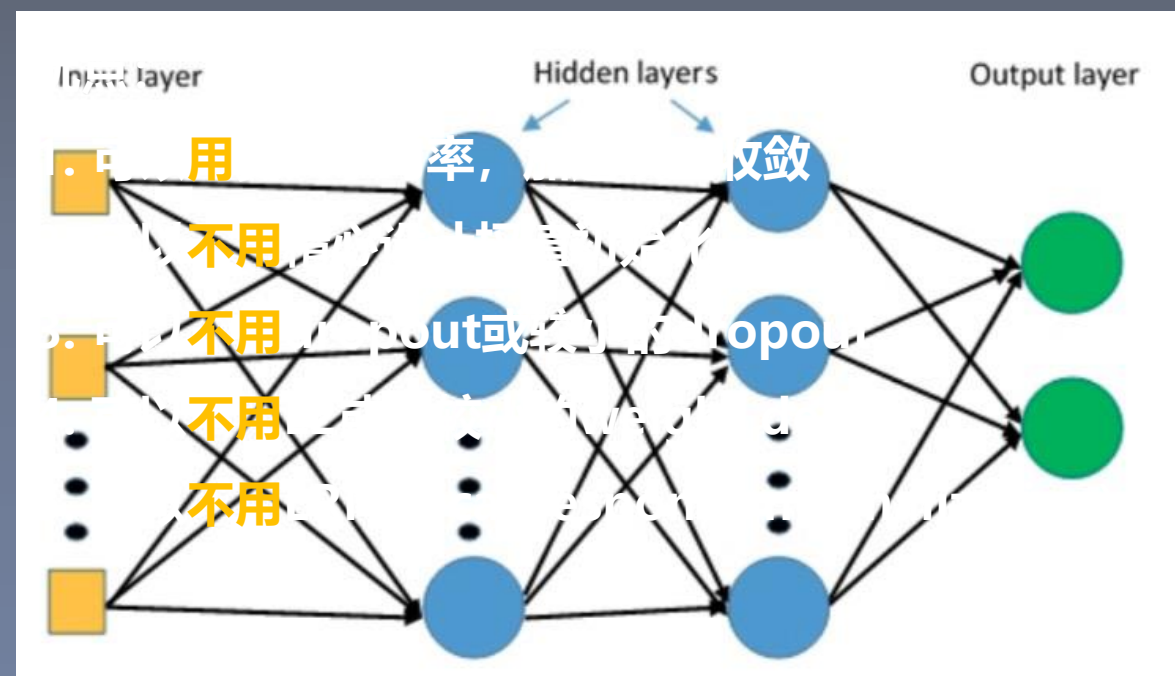
$$H_{11} = \sum_{i=0}^n X_i * W_{1i}$$

$$\begin{aligned} D(H_{11}) &= \sum_{i=0}^n D(X_i) * D(W_{1i}) \\ &= n * (1 * 1) \\ &= n \end{aligned}$$

$$\text{std}(H_{11}) = \sqrt{D(H_{11})} = \sqrt{n}$$

$$D(H_1) = n * D(X) * D(W) = 1$$

$$D(W) = \frac{1}{n} \Rightarrow \text{std}(W) = \sqrt{\frac{1}{n}}$$



$X$   $W_1$   $H_1$   $W_2$   $H_2$   $W_3$   $out$

# Batch Normalization

## Batch Normalization

### \_BatchNorm

- `nn.BatchNorm1d`
- `nn.BatchNorm2d`
- `nn.BatchNorm3d`

参数:

- `num_features`: 一个样本特征数量 (最重要)
- `eps`: 分母修正项
- `momentum`: 指数加权平均估计当前mean/var
- `affine`: 是否需要affine transform
- `track_running_stats`: 是训练状态, 还是测试状态

```
__init__(self, num_features,  
eps=1e-5,  
momentum=0.1,  
affine=True,  
track_running_stats=True)
```

# Batch Normalization

## Batch Normalization

- `nn.BatchNorm1d`
- `nn.BatchNorm2d`
- `nn.BatchNorm3d`

主要属性：

- `running_mean`：均值
- `running_var`：方差
- `weight`：affine transform中的gamma
- `bias`：affine transform中的beta

$$\hat{x}_i \leftarrow \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}}$$

$$y_i \leftarrow \gamma \hat{x}_i + \beta \equiv \text{BN}_{\gamma, \beta}(x_i)$$

训练：均值和方差采用指数加权平均计算  
测试：当前统计值

`running_mean = (1 - momentum) * pre_running_mean + momentum * mean_t`

`running_var = (1 - momentum) * pre_running_var + momentum * var_t`

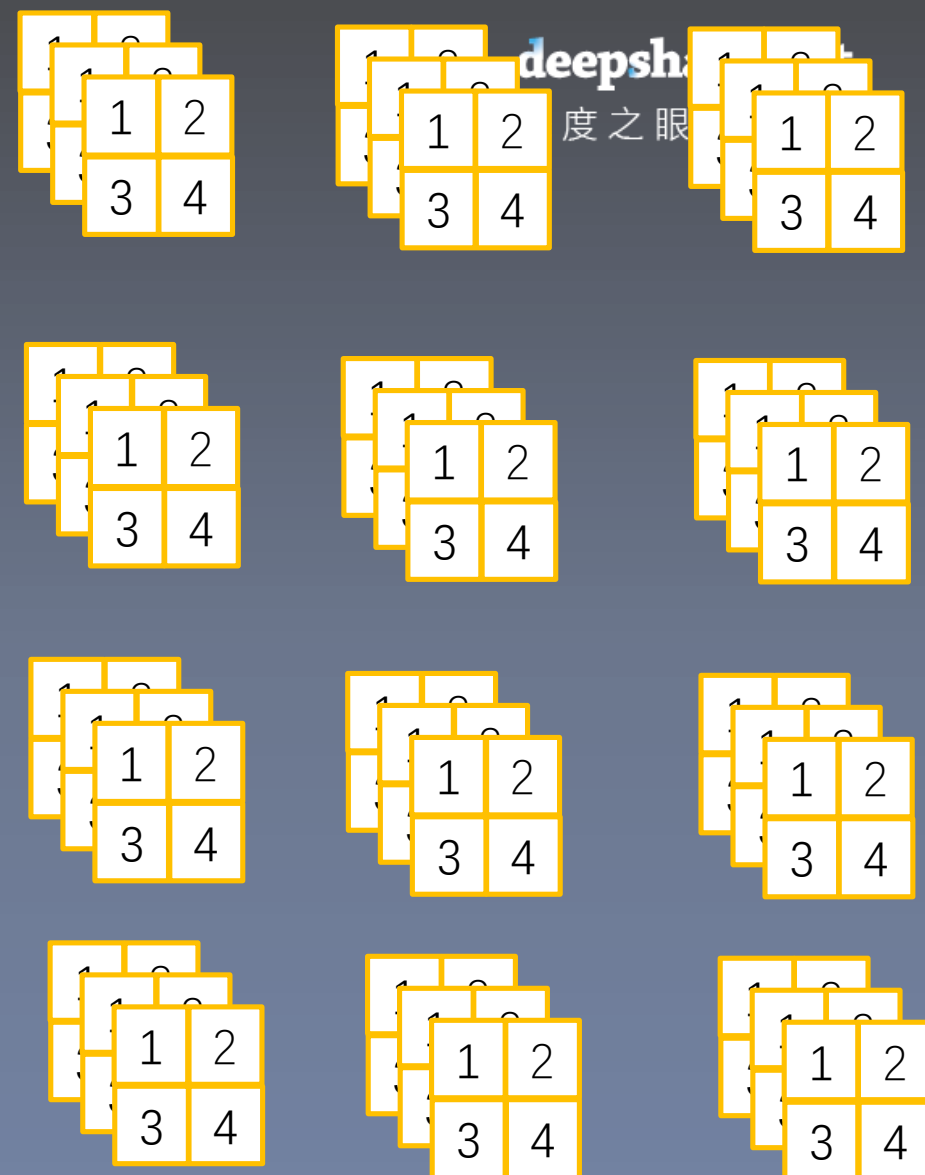
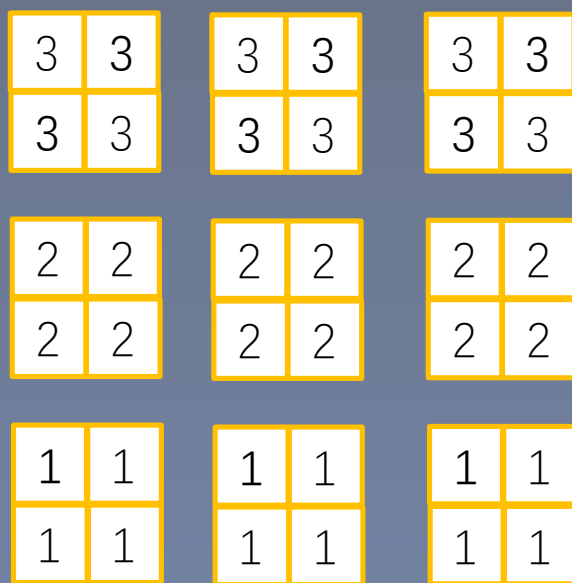
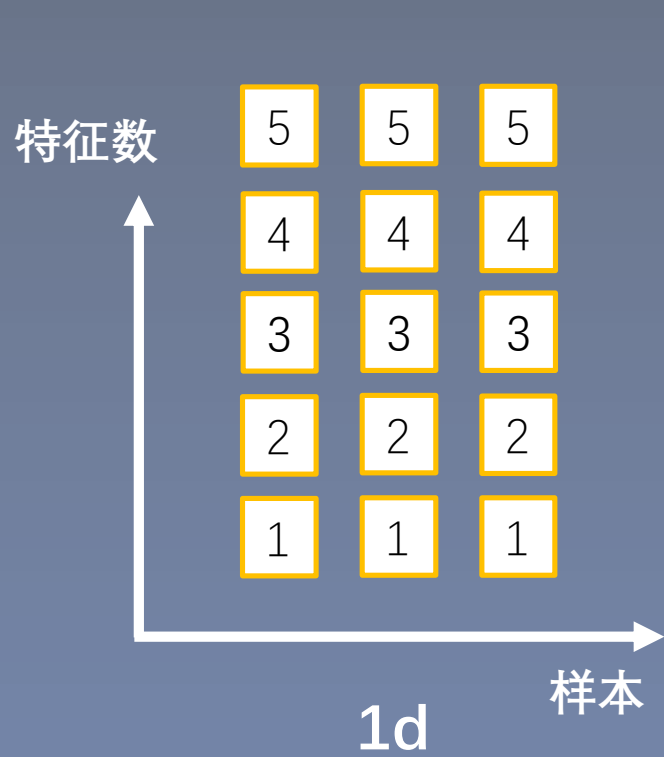
关注公众号深度之眼，后台回复论文，获取60篇AI必读经典前沿论文



# Batch Normalization

## Batch Normalization

- `nn.BatchNorm1d` input=  $B \times \text{特征数} \times 1\text{d}$  特征
- `nn.BatchNorm2d` input=  $B \times \text{特征数} \times 2\text{d}$  特征
- `nn.BatchNorm3d` input=  $B \times \text{特征数} \times 3\text{d}$  特征

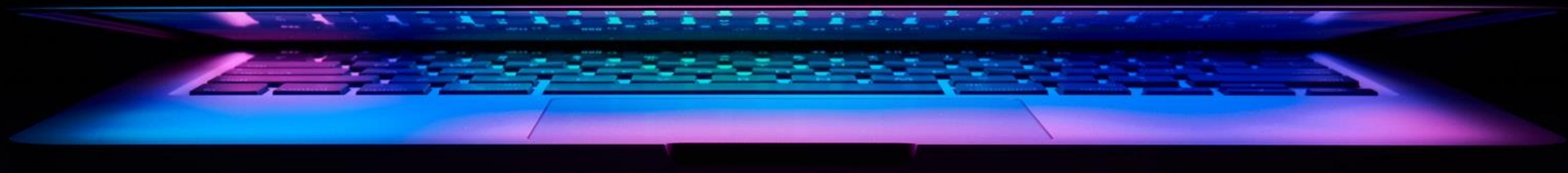


# —— 结 语 ——

在这次课程中，学习了Batch Normalization

在下次课程中，我们将会学习

**其它 Normalization**



关注公众号深度之眼，后台回复论文，获取60篇AI必读经典前沿论文



**deepshare.net**

深度之眼

联系我们：

电话：18001992849

邮箱：[service@deepshare.net](mailto:service@deepshare.net)

QQ：2677693114



公众号



客服微信

关注公众号深度之眼，后台回复论文，获取60篇AI必读经典前沿论文