

Input

MNIST 数据集读入

特征→矩阵、标签→one hot 向量

构建网络模型 (Model Layers)

开始模型训练 $k := 0$

打乱训练集

将训练集分为多个大小为B的 Mini-batch

取出一个batch

一个batch中的样本

$(\mathbf{x}^{(1)}, \hat{\mathbf{y}}^{(1)})$

\vdots

$(\mathbf{x}^{(B)}, \hat{\mathbf{y}}^{(B)})$

多线程优化

计算梯度
propagation

\vdots

计算梯度
propagation

梯度平均值
 $\nabla \mathcal{L}(\mathbf{W}, \mathbf{b})$

更新网络参数

取出下一个batch

完成一次训练
 $k := k + 1$

$k < \text{echo}$

False

使用验证集, 评估泛用性
evaluate

$(\mathbf{x}^{(i)}, \hat{\mathbf{y}}^{(i)})$

网络计算
propagation

判断预测值与真实
值是否相等

输出正确率

END

True

取下一个样本

Model 类

echo : 训练次数

batch-size : 每个 Mini-batch 大小, 下面记为B

loss function : 损失函数 $\mathcal{L}(\mathbf{y}, \hat{\mathbf{y}})$

layers : 网络结构, 以 Layer 类为元素的列表

Layer 类

n : 前一层神经元个数

m : 当前层神经元个数

activation function : 激活函数

W : 权重矩阵 ($\mathbb{R}^{m \times n}$)

b : 偏置 ($\mathbb{R}^{m \times 1}$)

z : 净活性值 $\mathbf{W}\mathbf{x} + \mathbf{b}$