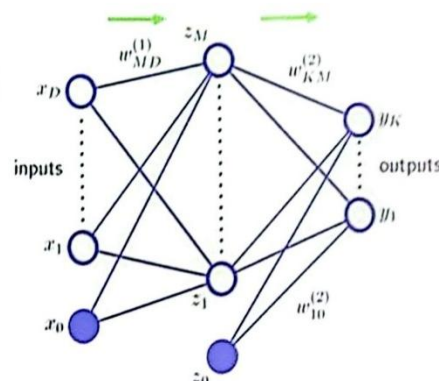


1. 已知某神经网络如右图所示，输入层、隐藏层与输出层变量维度分别为D、M和K维。隐藏层的激活函数为Sigmoid函数 $\sigma(a)$ ，输出层为线性映射，使用平方和损失函数。



- (1) 试计算该神经网络的参数数量；
- (2) 试逐层写出该神经网络的前向传递函数，注意 $x_0 = 1$ 与 $z_0 = 1$ 用于表示偏置；
- (3) 两层网络参数分别记为 w_{μ} 与 w_{kj} ，试求 $\frac{\partial E_n}{\partial w_{\mu}}$ 。

2. 将以下术语翻译为英文

- (1) 神经网络
- (2) 随机梯度下降
- (3) 线性回归
- (4) 逻辑回归
- (5) 交叉熵损失函数

3. 简要回答以下问题

- (1) 什么是机器学习？
- (2) 在机器学习模型中引入 L_2 正则化项的目的是什么？其工作原理是什么？
- (3) 使用数值差分方法计算神经网络损失函数的梯度时，其计算复杂度如何？请提供推理过程。

1、某二分类器在测试集上共有 1000 个样本，其中：正类样本 400 个，被正确分类为正类 320 个；负类样本 600 个，被误判为正类 90 个。

[1] 计算准确率、错误率；

[2] 计算查准率 P、查全率 R 以及 F1 值。

2、比较 Filter、Wrapper、Embedded 三类特征选择方法的主要区别，并各举一个典型算法或模型。

3、简要说明 PCA 的两种等价视角，为什么这两种视角得到的是同一组主成分？

4、考虑链式结构 $A \rightarrow B \rightarrow C$ 。给定： $p(A=1)=0.4$, $p(B=1 | A=1)=0.8$, $p(B=1 | A=0)=0.1$, $p(C=1 | B=1)=0.9$, $p(C=1 | B=0)=0.2$ 。

[1] 计算边缘概率 $p(C=1)$ 。

5、假设输入是一张 $224 \times 224 \times 3$ 的彩色图像。

- 方案 A：直接接一层全连接层，输出 1000 维分类 logits。
- 方案 B：先接一个 2D 卷积层：kernel size = 7×7 、stride=2、Zero padding，输出通道数 $C_0=64$ 。然后将该特征图展平，再接全连接层输出 1000 维 logits。

假设卷积后特征图大小为 $112 \times 112 \times 64$ （与常见设置类似）。

[1] 计算方案 A 的参数数量。

[2] 计算卷积层自身参数数量。

[3] 计算方案 B 中最终全连接层的参数量，并比较 A 与 B 总参数量数量级的差异。

一、令线性回归模型为

$$y = \mathbf{w}^T \mathbf{x} + b$$

1. 试给出优化该模型参数的均方误差损失函数 $E(\mathbf{w}, b)$;
2. 试给出梯度下降算中权重 \mathbf{w} 和偏差 b 的更新公式。
3. 试给出包含关于权重的 L2 正则化的优化目标函数，并求权重 \mathbf{w} 与偏差 b 的更新公式。

二、试给出 Sigmoid 函数 $\sigma(z)$ 的数学表达式，并求 $\frac{d\sigma}{dz}$ 。

三、试分析训练获得的线性回归模型可能的误差来源。