

## 机器学习与深度学习习题集（中）

这是机器学习与深度学习习题的第二部分，为《机器学习-原理，算法与应用》一书编写，二者配合使用。习题集的绝大部分题目都可以在此书中找到答案。同时也可以用作高校相关专业的机器学习，深度学习课程习题集。后续我们将给出最后一部分，以及整个习题集的完整答案。

### 第 13 章 Boosting 算法

1. 写出 AdaBoost 算法强分类器的预测公式。
2. 写出 AdaBoost 的训练算法。
3. 证明 AdaBoost 强分类器在训练样本集上的错误率上界是每一轮调整样本权重时权重归一化因子  $Z_t$  的乘积，即下面的不等式成立

$$p_{error} = \frac{1}{l} \sum_{i=1}^l \mathbb{I}[\text{sgn}(F(\mathbf{x}_i)) \neq y_i] \leq \prod_{t=1}^T Z_t$$

4. 接第 3 题，假设  $e_t$  为第  $t$  个弱分类器的错误率。证明下面的不等式成立

$$\prod_{t=1}^T Z_t = \prod_{t=1}^T 2\sqrt{e_t(1-e_t)} = \prod_{t=1}^T \sqrt{(1-4\gamma_t^2)} \leq \exp\left(-2 \sum_{t=1}^T \gamma_t^2\right)$$

其中

$$\gamma_t = \frac{1}{2} - e_t$$

5. 简述广义加法模型的原理。
6. 离散型 AdaBoost 的损失函数是什么函数？
7. 从广义加法模型和指数损失函数推导 AdaBoost 的训练算法。
8. 解释实数型 AdaBoost 算法的原理。
9. AdaBoost 算法的弱分类器应该如何选择？

10.简述梯度提升算法的原理。

11.假设使用均方误差函数，梯度提升算法如何解决回归问题？

12.梯度提升算法如何解决二分类问题？

13.对于多分类问题，梯度提升算法的预测函数是  $F_k(\mathbf{x})$ 。样本属于每个类的概率为：

$$p_k(\mathbf{x}) = \frac{\exp(F_k(\mathbf{x}))}{\sum_{l=1}^K \exp(F_l(\mathbf{x}))}$$

如果加上限制条件

$$\sum_{l=1}^K F_l(\mathbf{x}) = 0$$

证明如下结论成立

$$F_k(\mathbf{x}) = \ln p_k(\mathbf{x}) - \frac{1}{K} \sum_{l=1}^K \ln p_l(\mathbf{x})$$

14.解释 XGBoost 算法的原理。

15.XGBoost 算法为何要泰勒展开到二阶？

16.比较随机森林与 AdaBoost 算法的异同。

17.比较 AdaBoost 算法与梯度提升算法的异同。

18.随机森林为什么能够减小方差？

19.AdaBoost 算法为什么能够降低偏差？

## 第 14 章 深度学习概论

1.列举你所知道的深度学习的典型应用。

2.列举你所知道的深度神经网络结构。

3.列举你所知道的深度生成模型。

4.列举你所知道的深度强化学习算法。

- 5.解释深度强化学习算法的原理。
- 6.人工特征存在哪些问题?
- 7.什么是维数灾难?
- 8.解释自动编码器的原理, 写出自动编码器训练时的目标函数。
- 9.解释去噪自动编码器的原理。
- 10.解释稀疏自动编码器的原理, 写出它训练时的目标函数。
- 11.解释收缩自动编码器的原理。
- 12.写出玻尔兹曼分布的概率质量函数。
- 13.解释二部图的定义。
- 14.简述 RBM 的网络结构。
- 15.推导 RBM 隐变量和可见变量条件概率的计算公式。
- 16.写出 RBM 的训练目标函数。
- 17.简述 DBN 的原理。
- 18.简述 DBN 的原理。
- 19.比较 DBN 与 DBM 的异同。

## 第 15 章 卷积神经网络

- 1.解释卷积运算的原理。
- 2.假设输入图像为

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 0 & 1 & 1 & 1 \\ 2 & 3 & 4 & 5 \end{bmatrix}$$

卷积核为

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

假设卷积步长为 1，计算卷积结果。

假设卷积步长为 2，计算卷积结果。

3.解释多通道卷积的原理。

4.解释 1x1 卷积的原理，这种卷积有什么用途？

5.解释反卷积（转置卷积）的原理，这种卷积有什么用途？

6.假设输入图像为

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 0 & 1 & 1 & 1 \\ 2 & 3 & 4 & 5 \end{bmatrix}$$

卷积核为

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

计算反卷积的结果。

7.解释空洞卷积的原理，并举例说明。

8.解释可变形卷积的原理。

9.卷积层为什么需要激活函数？

10.解释池化运算的原理，为什么需要池化运算？

11.比较均值池化与最大值池化。

12.假设输入图像为

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 0 & 1 & 1 & 1 \\ 2 & 3 & 4 & 5 \end{bmatrix}$$

池化核为 2x2

计算步长为 1 时均值池化的结果，步长为 2 时均值池化的结果。

计算步长为 1 时最大值池化的结果，步长为 2 时最大值池化的结果。

13.解释卷积运算填充（padding）的原理，这种方法有什么用？

14.推导卷积层反向传播时的参数梯度计算公式。

- 15.推导卷积层反向传播时的误差项递推计算公式。
- 16.推导均值池化和最大值池化在反向传播时的误差项递推计算公式。
- 17.解释 mini-batch 随机梯度下降法的原理。
- 18.列举你所知道的神经网络权重初始化算法。
- 19.列举你所知道的神经网络学习率计算算法。
- 20.列出 AlexNet 的主要创新点。
- 21.解释 VGG 网络的主要创新点。
- 22.解释 GoogLeNet 网络 Inception 模块的原理。
- 23.解释残差网络的原理。
- 24.简述 DenseNet 的原理。
- 25.简述用矩阵乘法实现卷积运算的原理。如果输入图像为三通道

$$\begin{bmatrix} 1 & 0 & 1 & 1 \\ 2 & 2 & 3 & 0 \\ 3 & 5 & 1 & 2 \\ 1 & 0 & 2 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 2 & 3 & 0 \\ 2 & 3 & 5 & 2 \\ 1 & 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 & 1 \\ 2 & 4 & 3 & 0 \\ 3 & 5 & 3 & 0 \\ 1 & 0 & 2 & 0 \end{bmatrix}$$

卷积核为

$$\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

用矩阵乘法实现卷积运算。

- 26.softmax 交叉熵的计算公式为

$$y_i^* = \frac{\exp(x_i)}{\sum_{j=1}^k \exp(x_j)} \quad L = -\frac{1}{n} \sum_{i=1}^n y_i^T \log y_i^*$$

推导损失函数对输入值的梯度计算公式。

- 27.对于离散型概率分布，证明当两个概率分布相等时交叉熵具有极值，且为极小值。
- 28.对于分类问题，为什么通常使用交叉熵而不使用欧氏距离损失？

- 29.简述 AdaGrad 算法的原理，该算法存在什么问题？
- 30.简述 AdaDelta 算法的原理。
- 31.简述 Adam 算法的原理。
- 32.什么是激活函数的饱和性？
- 33.简述 Batch Normalization 的原理，这种技术解决了什么问题？
- 34.简述 Drop out 极值的原理，该机制在训练和预测时是如何实现的？
- 35.简述 early stop 的原理，这种技术解决了什么问题？

## 第 16 章 循环神经网络

- 1.循环神经网络是如何具有记忆功能的？
- 2.写出循环层的计算公式。
- 3.如果输出层采用 softmax 交叉熵，推导具有一个输入层，一个输出层，一个循环层的 RNN 的反向传播算法计算公式。
- 4.为什么 RNN 学习长期依赖存在困难？
- 5.简述 LSTM 的原理。
- 6.简述 GRU 的原理。
- 7.简述双向循环神经网络的原理。
- 8.简述 CTC 的原理。
- 9.简述 CTC 中 B 函数的工作原理。
- 10.写出 CTC 训练时的目标函数。
- 11.推导 CTC 的解码算法。
- 12.用实例解释 beam search 的原理。

13.简述 seq2seq 的原理。

14.seq2seq 是如何解决机器翻译问题的？

15.RNN 是如何解决中文分词问题的？

## 第 17 章 生成对抗网络

1.如何生成均匀分布的随机数？

2.如何生成正态分布的随机数？

3.解释拒绝采样算法的原理。

4.解释 MCMC 采样算法的原理。

5.解释 M-H 采样算法的原理。

6.解释 Gibbs 采样算法的原理。

7.简述 GAN 的原理。

8.写出 GAN 的训练目标函数。

9.写出 GAN 的训练算法。

10.GAN 训练时的目标函数为

$$\min_G \max_D V(D, G) = \mathbb{E}_{\mathbf{x} \sim p_{data}(\mathbf{x})} [\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim p_z(\mathbf{z})} [\log (1 - D(G(\mathbf{z})))]$$

证明：如果生成模型固定不变，使得目标函数取得最优值的判别模型为

$$D_G^*(\mathbf{x}) = \frac{p_{data}(\mathbf{x})}{p_{data}(\mathbf{x}) + p_g(\mathbf{x})}$$

证明：当且仅当

$$p_g = p_{data}$$

时这个目标函数取得最小值，且最小值为  $-\log 4$ 。

11.证明 KL 散度是非负的。

12.简述 CGAN 的原理。

13.什么是模型坍塌问题？

14.简述 WGAN 的原理，写出其训练目标函数。

15.GAN 的训练目标函数与 logistic 回归有何不同？

16.简述 Inception Score 的原理。

17.简述 Frechet Inception Distance 的原理。

## 第 18 章 聚类算法

1.列举常见的聚类算法类型。

2.解释层次聚类算法的原理。

3.层次聚类算法的实现依赖于计算簇之间的距离，有哪些距离定义？

4.写出 k 均值算法的优化目标函数。

5.写出 k 均值算法的流程。

6.k 均值算法的簇中心如何初始化？

7.如何确定 k 均值算法的 k 值？

8.写出高斯混合模型的概率密度函数。

9.写出高斯混合模型的对数似然函数。

10.高斯混合模型的权重系数之和为什么必须为 1？

11.如果直接求解高斯混合模型对数似然函数的极值，存在什么问题？

12.写出 Jensen 不等式。

13.证明 Jensen 不等式成立。

14.Jensen 不等式在什么时候取等号。

15.简述 EM 算法的原理。

16.写出 EM 算法的流程。



17.证明 EM 算法收敛。

18. 用 EM 算法求解高斯混合模型。

19.简述 DBSCAN 算法的原理。

20.简述 OPTICS 算法的原理。

21.OPTICS 算法对 DBSCAN 算法的主要改进是什么？

22.OPTICS 算法是如何根据样本的排序得到聚类结果的？

23.简述核密度估计的原理。

24.简述 mean shift 算法的原理。

25.推导 mean shift 的梯度计算公式

$$\mathbf{m} = \frac{\sum_{i=1}^n g(\mathbf{x}_i - \mathbf{x}) \mathbf{x}_i}{\sum_{i=1}^n g(\mathbf{x}_i - \mathbf{x})} - \mathbf{x}$$

其中

$$g(x) = -k'(x)$$

$$K\left(\frac{\mathbf{x} - \mathbf{x}_i}{h}\right) = ck\left(\left\|\frac{\mathbf{x} - \mathbf{x}_i}{h}\right\|^2\right)$$

26.什么是外部指标？列举常见的外部指标。

27.什么是内部指标？列举常见的内部指标。

28.简述谱聚类算法的原理。