

# 参考答案

## 机器学习与深度学习理论基础

### 单项选择题答案：

1. **B. 监督学习、无监督学习、强化学习**
  - 解释：这是机器学习最经典和基础的三大分类。监督学习从标记数据中学习，无监督学习从未标记数据中发现模式，强化学习通过与环境交互学习策略。
2. **A. 最小化模型在训练数据上的预测误差，通常通过最小化损失函数（如均方误差）来实现**
  - 解释：梯度下降是一种优化算法，通过迭代调整模型参数以逐步减小损失函数的值，从而使模型预测更接近真实值。
3. **B. 将数据集划分为具有相似特征的K个不同的簇**
  - 解释：K-means的核心思想是将数据点分配到K个簇中，使得同一簇内的数据点尽可能相似，不同簇之间的数据点尽可能不同。
4. **B. 为网络引入非线性，使其能够学习和表示复杂的模式和数据关系**
  - 解释：如果没有激活函数或使用线性激活函数，多层神经网络本质上等同于一个单层线性模型，无法学习复杂函数。非线性激活函数是深度学习模型强大表达能力的关键。
5. **B. 从网络输出层开始，将误差逐层向后传播，并根据误差梯度更新网络权重**
  - 解释：反向传播算法利用链式法则计算损失函数相对于网络各层权重的梯度，然后使用这些梯度通过优化算法（如梯度下降）更新权重。
6. **C. 通过可学习的滤波器（卷积核）提取输入数据的局部特征，如边缘、纹理等**
  - 解释：卷积操作通过在输入上滑动卷积核并计算点积，来响应特定的局部模式，这些模式（如边缘、角点）是构成更复杂视觉特征的基础。
7. **B. 使得非常深的网络也能有效训练，缓解了梯度消失和网络退化问题**
  - 解释：残差连接允许梯度更直接地反向传播到较早的层，从而缓解了深度网络中常见的梯度消失问题，并解决了网络层数增加时可能出现的性能退化问题。
8. **C. 难以捕捉和学习序列中的长期依赖关系，容易出现梯度消失或爆炸**
  - 解释：LSTM通过其门控机制（输入门、遗忘门、输出门）和细胞状态，能够有选择地记忆和遗忘信息，从而更有效地处理长序列中的依赖关系。
9. **B. 提供了大量预训练的Transformer模型和便捷的API，极大地简化了SOTA模型的应用和微调**
  - 解释：Hugging Face Transformers通过提供易于使用的模型库、工具和社区，极大地推动了Transformer架构在NLP及其他领域的普及和发展。
10. **C. 正确地实现了权重衰减（Weight Decay）与L2正则化的解耦，使其更符合原始L2正则化的意图，通常能带来更好的泛化性能**
  - 解释：Adam中的L2正则化（权重衰减）与自适应学习率耦合，可能导致对具有较大梯度的权重施加较小的有效权重衰减。AdamW通过解耦这两者，使得权重衰减更像原始的L2正则化，通常在训练大型模型时效果更好。

### 11. B. 过拟合 (Overfitting)

- **解释：**过拟合指模型学习了训练数据中的噪声和细节，而不是潜在的数据分布，导致其在训练集上表现优异，但在新数据上泛化能力差。

### 12. B. 模型过于简单，未能捕捉到数据中的基本模式，导致在训练数据和测试数据上表现均不佳

- **解释：**高偏差通常与欠拟合相关，意味着模型对数据的假设过于简化，无法很好地拟合训练数据，因此在测试数据上表现也差。

### 13. C. 通过线性变换将高维数据投影到低维空间，同时最大程度地保留原始数据的方差

- **解释：**PCA寻找一组新的正交基（主成分），使得数据在这些基上的投影方差最大，从而在降维的同时保留数据的主要信息。

### 14. B. 将在一个任务上学到的知识（如特征、权重）应用于另一个相关但不同的任务，以提高学习效率和性能

- **解释：**迁移学习利用预训练模型在大规模数据集上学到的通用知识，将其作为新任务的起点，特别适用于目标任务数据量较少的情况。

### 15. B. 允许模型在处理输入序列或图像时，动态地将焦点放在与当前任务最相关的部分

- **解释：**注意力机制通过为输入的不同部分分配不同的权重，使模型能够“关注”对当前预测或决策更重要的信息，提高了模型的性能和可解释性。

## 多项选择题答案：

### 1. A. 根据房屋的面积、位置等特征预测其价格 B. 将新闻文章自动分类到体育、政治、科技等类别

- **解释：**A是回归问题（预测连续值），B是分类问题（预测离散类别），两者都使用带标签的数据进行训练，是监督学习的典型应用。C是聚类（无监督），D如果图像未标记则是无监督或自监督。

### 2. A. 聚类分析（如K-means） C. 降维（如PCA） D. 图像中物体边缘检测（通常作为特征提取，可用于后续监督或无监督任务，但边缘检测本身可视为一种模式发现）

- **解释：**A和C是经典的无监督学习任务。D边缘检测可以被视为一种低层次的特征提取，它在没有标签的情况下识别图像中的模式，可以用于后续的无监督或监督学习。B决策树分类是监督学习。

### 3. A. 输入层、隐藏层、输出层 B. 神经元之间的权重和偏置项 C. 每个神经元（除输入层外）的激活函数 D. 用于评估模型性能的损失函数和用于参数更新的优化算法

- **解释：**A, B, C是神经网络结构的核心组成部分。D是训练神经网络必不可少的过程和工具。

### 4. A. AlexNet C. ResNet (残差网络) D. VGG (Visual Geometry Group Network)

- **解释：**AlexNet, ResNet, VGG都是计算机视觉发展史上的里程碑式CNN架构。LSTM是RNN的一种，主要用于序列数据。

### 5. A. PyTorch C. TensorFlow

- **解释：**PyTorch和TensorFlow是目前深度学习领域最主流的两大开源框架。Scikit-learn主要用于传统机器学习。Apache Spark MLlib也支持一些机器学习，但不是专为深度学习设计，且不如前两者在深度学习社区流行。JAX也是一个高性能数值计算库，常用于研究和构建DL模型，可以认为是主流之一，但相较于Pytorch和Tensorflow，在工业界应用广泛性上可能稍逊。如果题目允许选多个，JAX也常被认为是主流。考虑到通常选择最核心的，A和C是最无争议的。