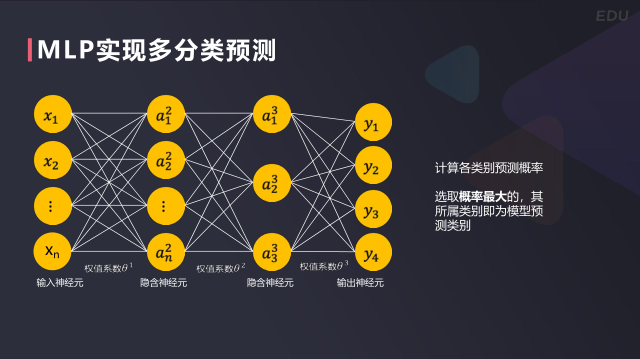
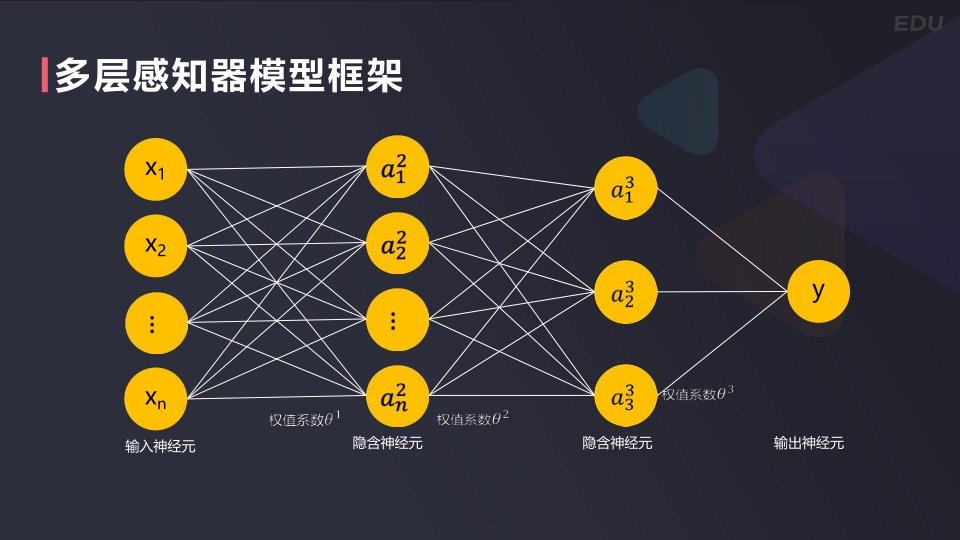
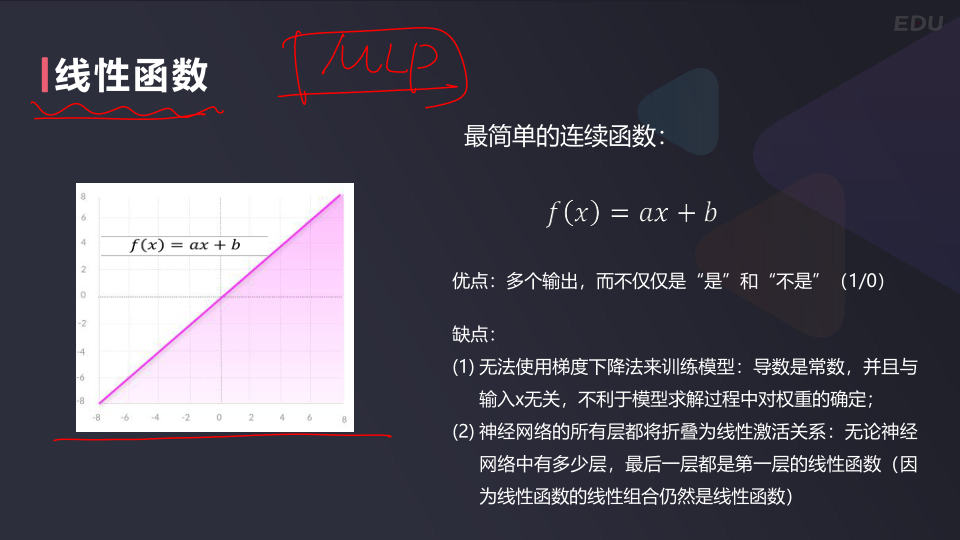
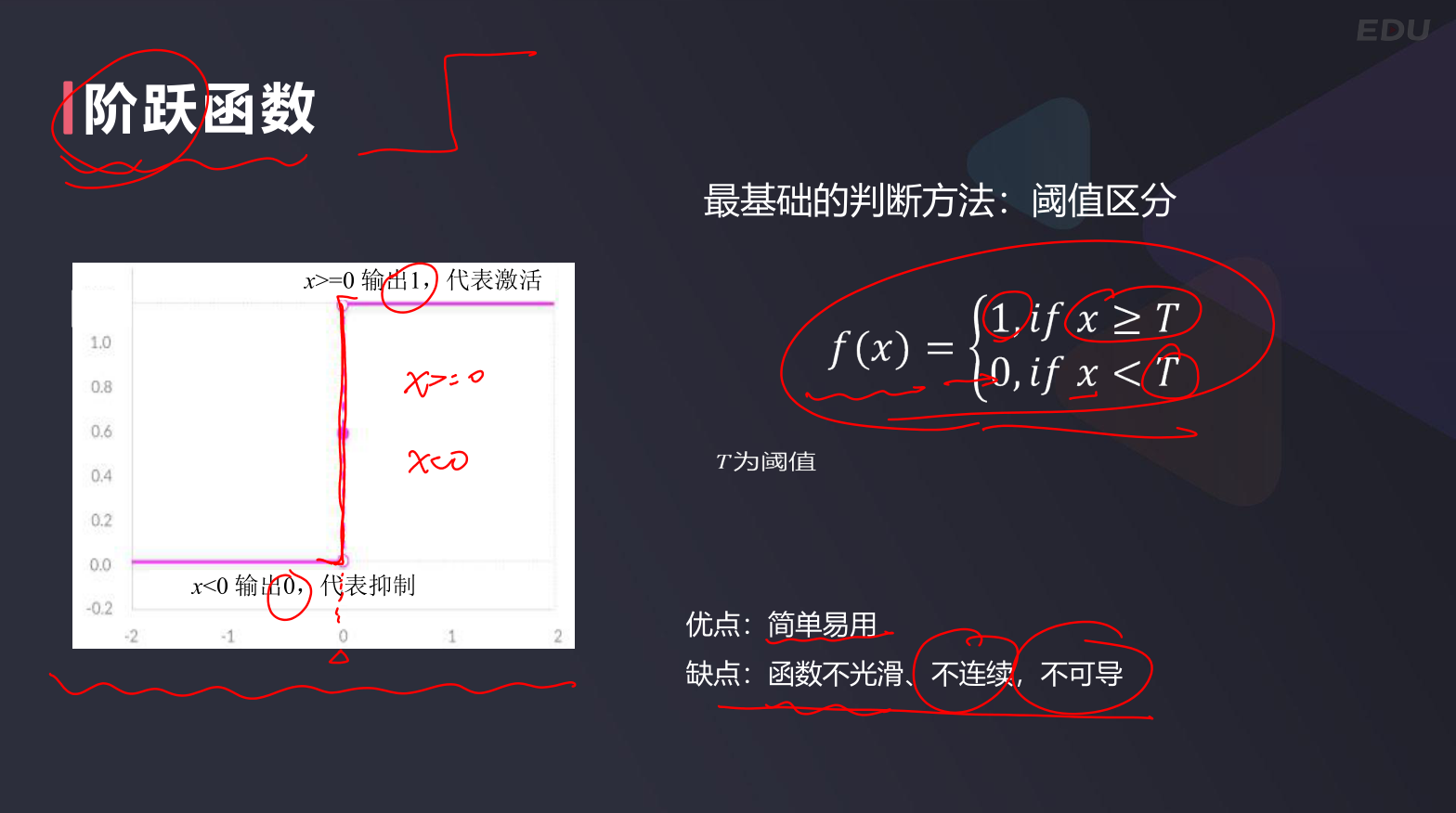
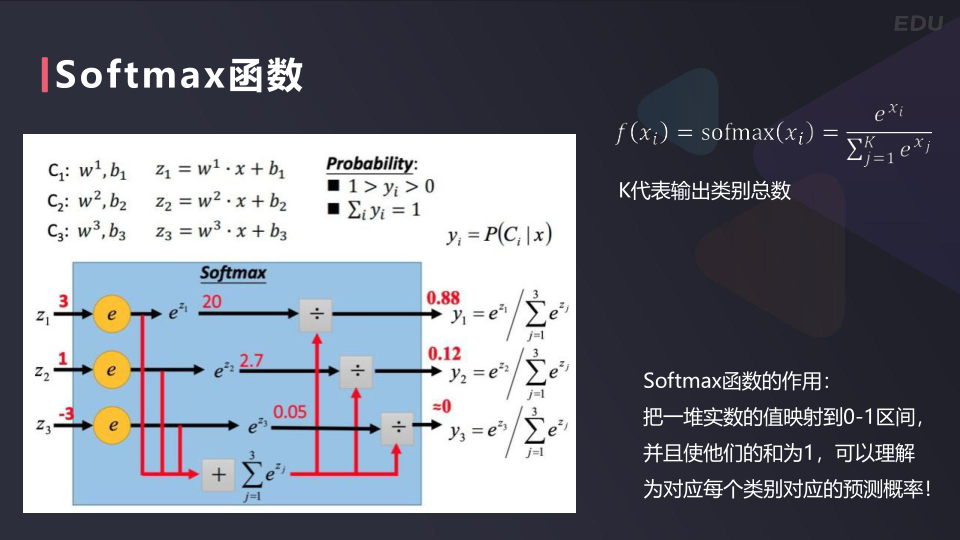
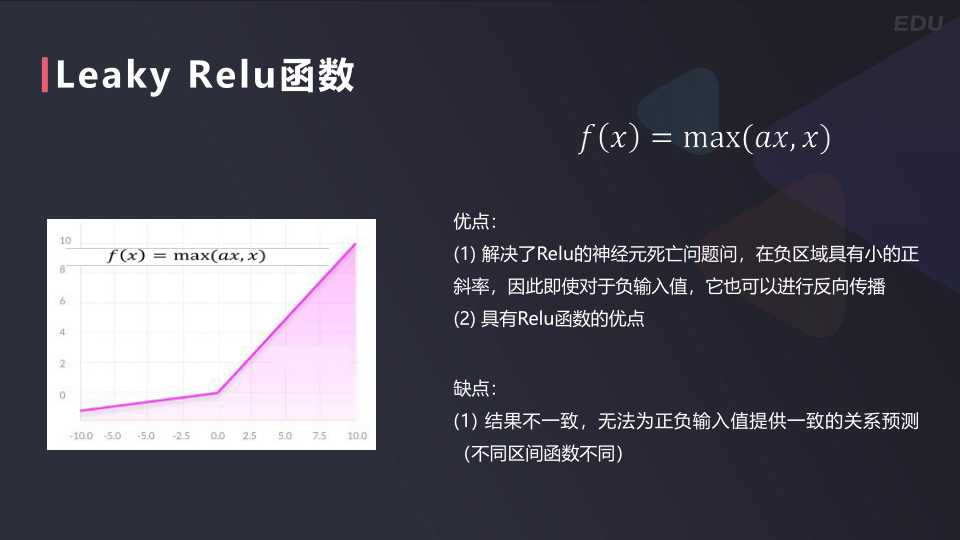
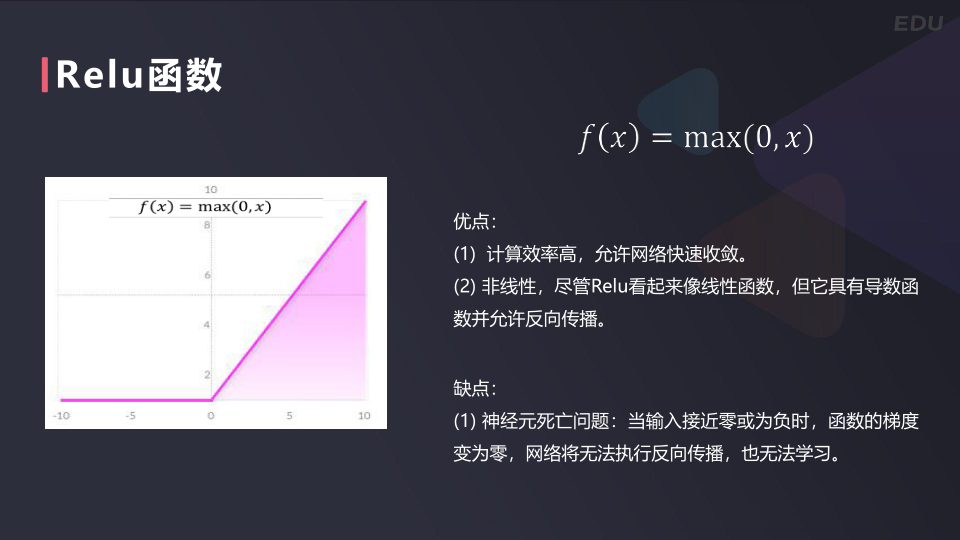
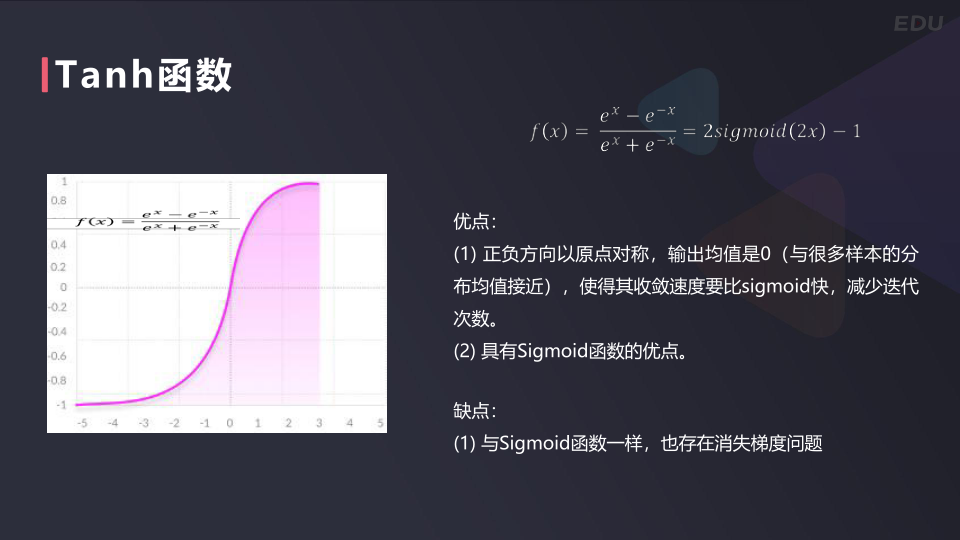
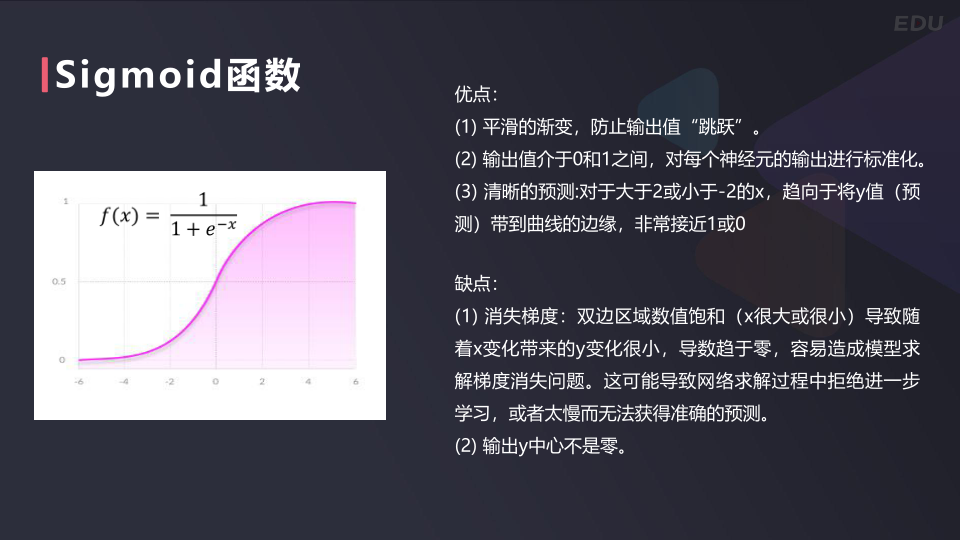
**人工神经网络（Artificial Neural Networks，ANN）**

**\*\*\*\*\*\*\*\*\*\*\*\* 多层感知器（MLP） \*\*\*\*\*\*\*\*\*\*\*\***



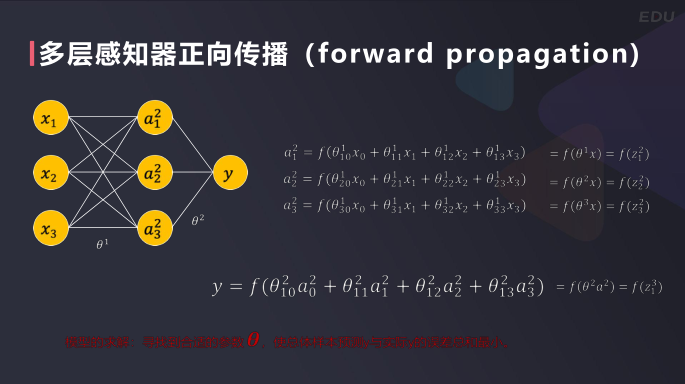
* **多层感知器（MLP）定义：又称为人工神经网络（ANN），一种类似于大脑神经突触联接的结构进行信息处理的数学模型。**
* **激活函数：在人工神经网络得神经元上运行得函数，负责将神经元得输入映射到输出端。激活函数主要类型包括：阶越函数、线性函数、Sigmoid函数、Tanh函数、Relu函数、Leaky Relu函数、Softmax函数。**





**根据场景决定使用合适的激活函数，没有最好的激活函数：**

* + **Sigmoid、Tanh：二分类任务输出层；模型隐藏层**
  + **Relu、Leaky Relu：回归任务；卷积神经网络隐藏层**
  + **Softmax：多分类任务输出层**
* **多层感知器正向传播：寻找到合适得参数θ，使总体样本预测y与实际y的误差总和最小，完成模型求解。**

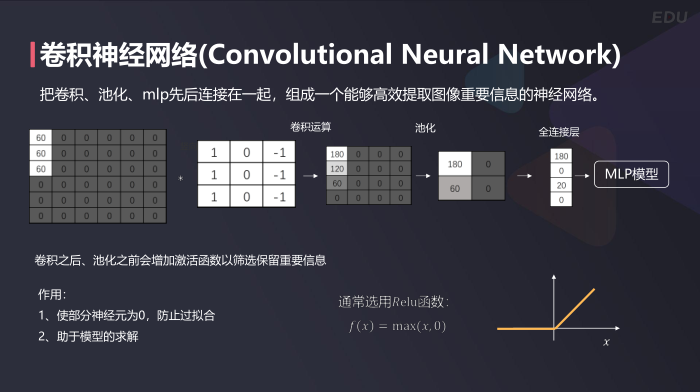
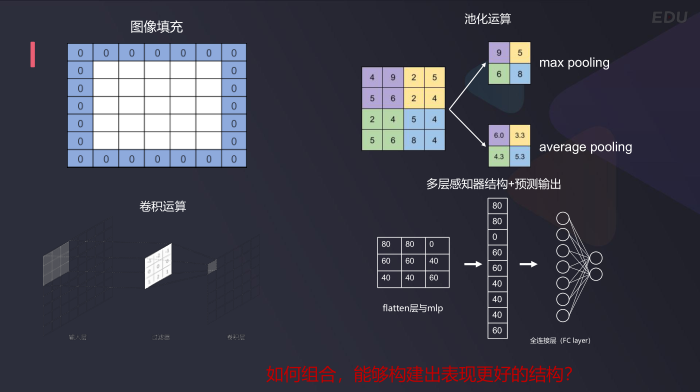


* **多层感知器反向传播：从后往前依次计算每层神经元的数值偏差，然后通过梯度下降法寻找到使偏差最小的参数θ，完成模型求解。**

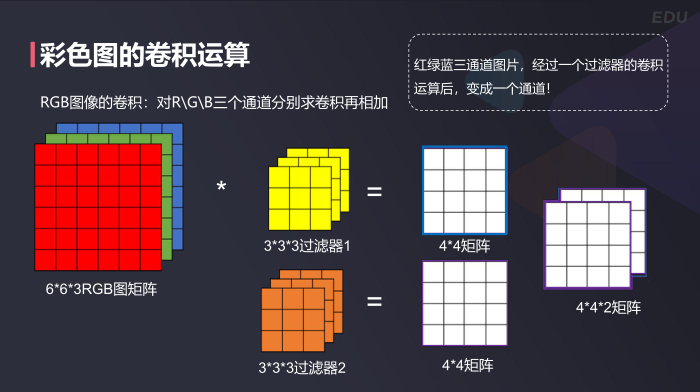
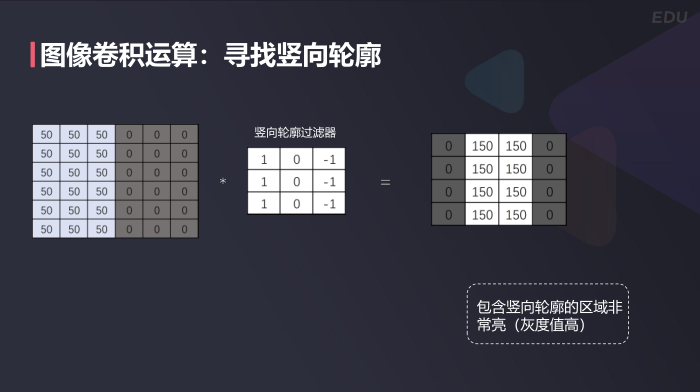
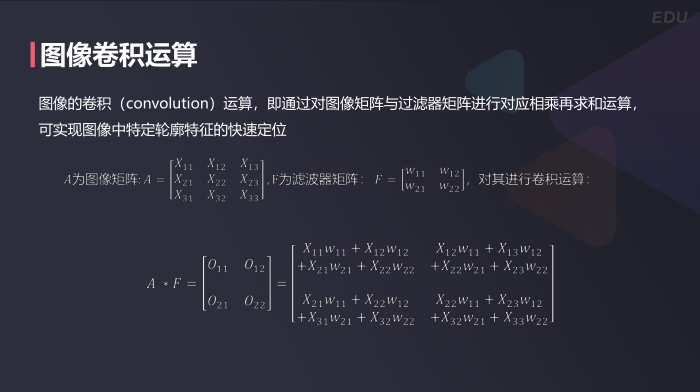


**\*\*\*\* 卷积神经网络（Convolutional Neural Network，CNN） \*\*\*\***

* **卷积神经网络：即把卷积、池化、MLP等先后连接在一起，组成的一个能够高效提取图像重要信息的神经网络。**
* **卷积神经网络两大特点：**
  1. **参数共享：同一特征过滤器可用于整张图片。**
  2. **稀疏连接：生成的特征图片每个节点只与原图片中特定节点连接。**
* **卷积神经网络的四大组成部分：卷积运算，池化，图像填充，MLP。**



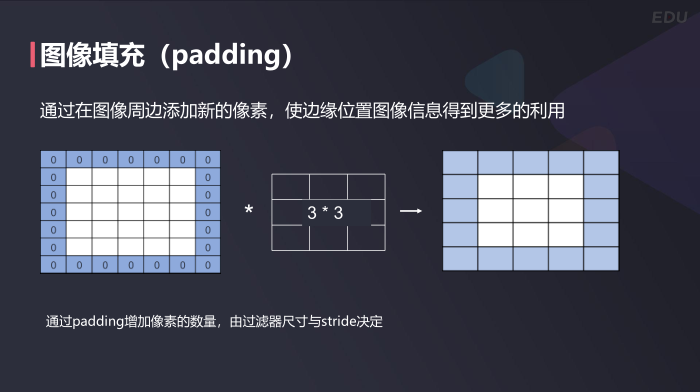
* **图像的卷积运算：即通过对图像矩阵与过滤器矩阵进行对应相乘再求和运算，可实现图像中特定轮廓特征的快速搜索。A与B的卷积通常表示为：A\*B或convolution（A, B）。**
  1. **需要卷积运算的原因：图像数据量过大，细节信息过多，通过先提取图像中的关键信息（轮廓）可以很大程度降低运算工作，因此先进行卷积。**
  2. **图像卷积运算的具体数学过程及灰度图像和RGB图像的不同卷积过程：**



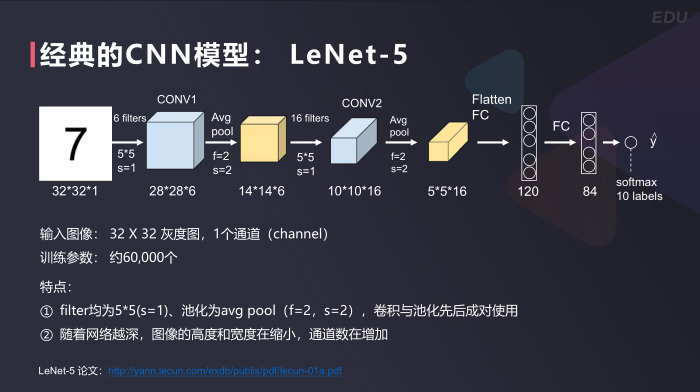
* 1. **过滤器：寻找合适的轮廓过滤器是卷积神经网络的核心，寻找合适的过滤器的过程本质即最小化输出损失函数的过程。轮廓过滤器包括竖向过滤器、横向过滤器、sobel过滤器等。（一个CNN中可以有多个过滤器）**
  2. **卷积计算后，需要增加激活函数以筛选保留重要信息。一般激活函数选择relu函数——可以使部分神经元为0，防止过拟合，并助于模型求解。**
* **池化：又称欠采样或下采样，即按照一定的规则对图像矩阵进行处理，实现信息压缩与数据降维。**
  1. **需要池化运算的原因：一张照片很多信息是不重要甚至重复多余的，使用所有像素点数据会增加运算量、可能导致过拟合、降低模型的容错性，因此需要使用池化，**
  2. **池化作用：特征降维，压缩数据和参数的数量，减小过拟合，同时提高模型的容错性。**
  3. **池化方法：最大法池化（Max-pooling）、平均法池化（Avg-pooling）等。同时还需要设置stride和pooling size，即窗口滑动步长和窗口大小。**



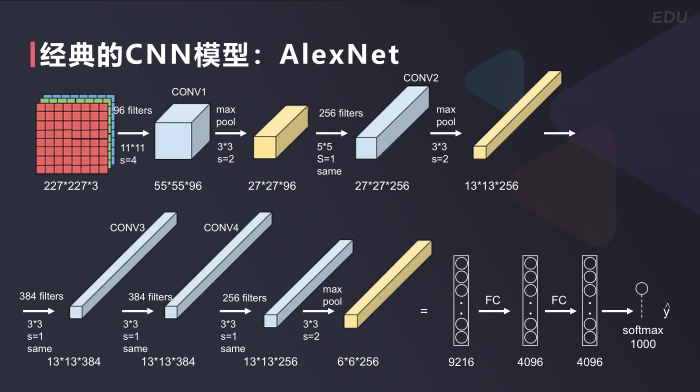
* **图像填充：在图像周边添加新的像素，使边缘位置图像信息得到更多的利用。**
  1. **需要图像填充的原因：使用卷积运算可以发现图像边缘信息的使用次数明显少于中间位置，容易被忽略 。**



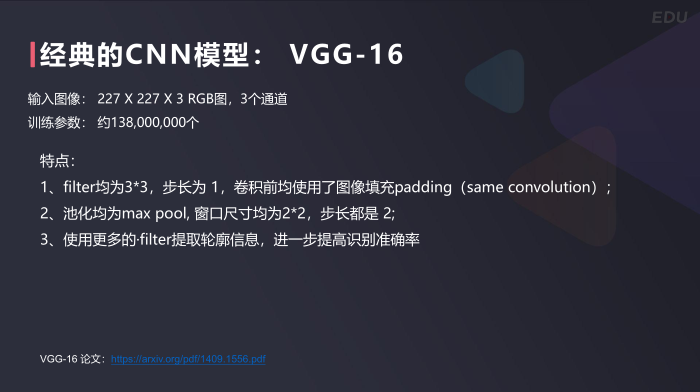
* **经典的CNN模型——LeNet-5**



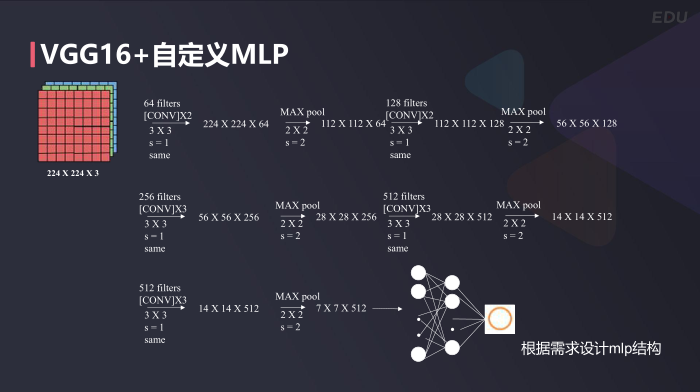
* **经典的CNN模型——AlexNet**



* **经典的CNN模型——VGG-16**

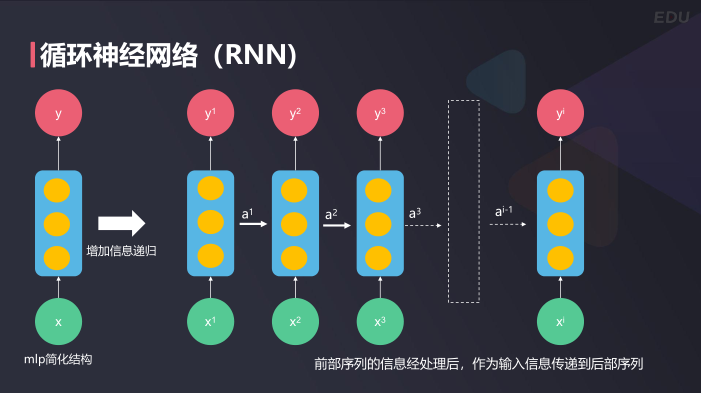


* **构建表现好的CNN的方法：**
  1. **考虑每一个步骤的参数配置：**
     1. **卷积：卷积算子filter尺寸，卷积算子filter数量等**
     2. **池化：池化方法，池化大小pooling size、卷积步长stride设置等**
     3. **填充：是否要填充，如何填充等**
  2. **参考经典的CNN结构，将其核心思想运用到新模型设计中。**
  3. **使用经典的CNN模型结构提取图像重要轮廓，再建立MLP模型。**
     1. **加载经典CNN模型，剥除其FC层，用于提取图像的重要轮廓信息**
     2. **把经过模型处理数据作为输入，分类结果作为输出，建立MLP模型**
     3. **模型训练，寻找图片不同类别对应的关键信息**

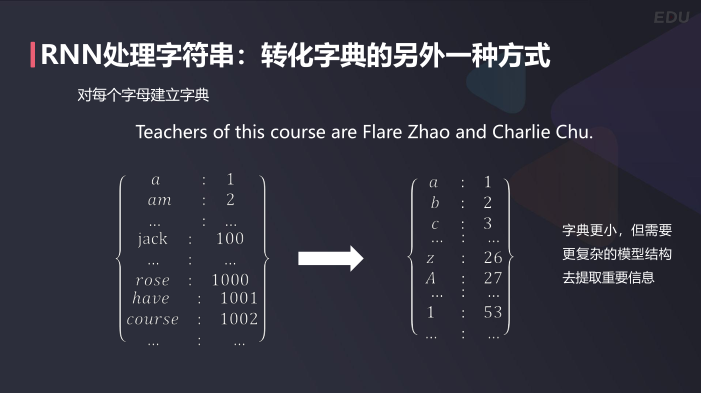
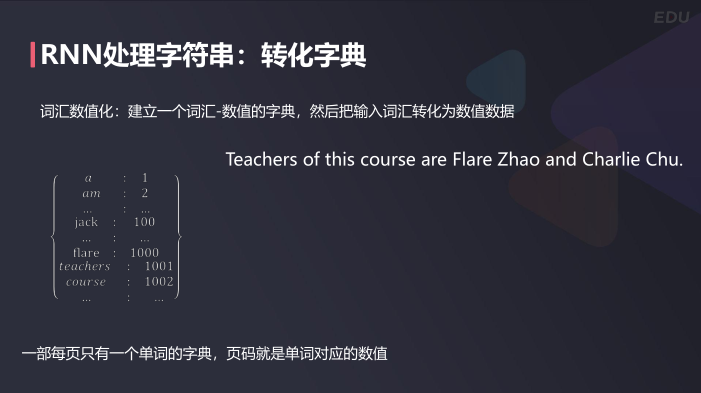


**\*\*\*\*\*\*\*\*\*\*\*\* 循环神经网络（RNN） \*\*\*\*\*\*\*\*\*\*\*\***

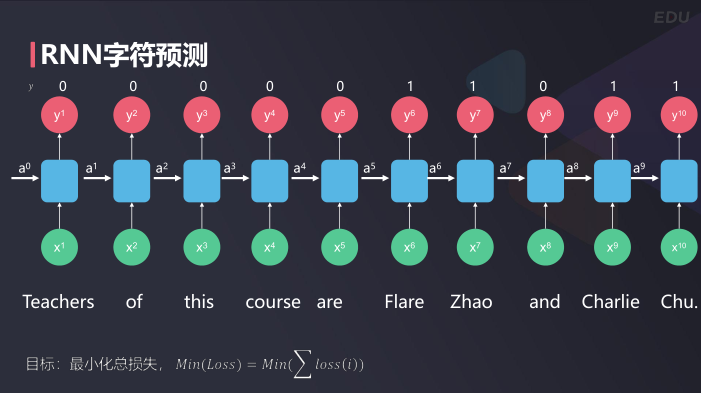
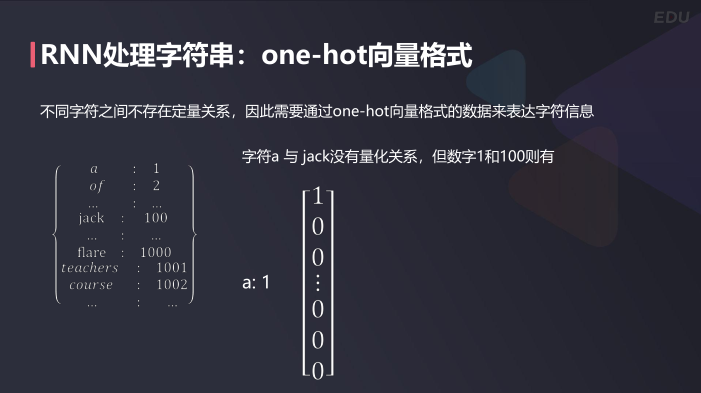
* **序列模型（Sequence Model）：输入或者输出中包含有序列数据的模型。**
* **序列模型两大特点：**
  1. **输入或输出元素之间是具有顺序关系。不同的顺序，得到的结果应该是不同的，比如“不睡觉”和“睡觉不”两个短语的意思是不同的。**
  2. **输入输出不定长。比如聊天机器人，聊天之间的对话长度都是不定的。**
* **序列模型的应用场景有：机器翻译、语音识别、行为预测、股价预测等。**
* **循环神经网络：以序列数据为输入，在序列的演进方向进行递归的神经网络。**



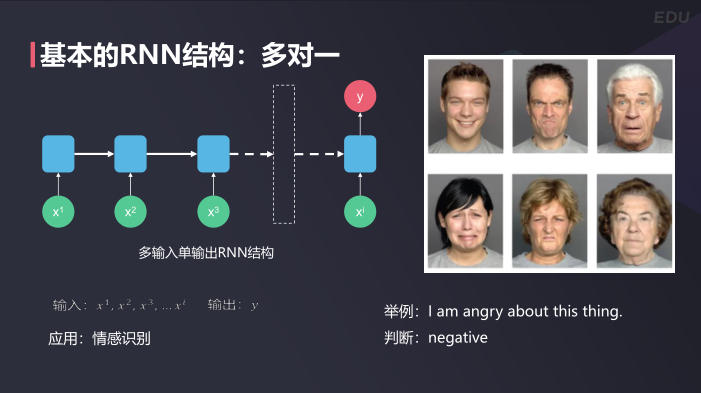
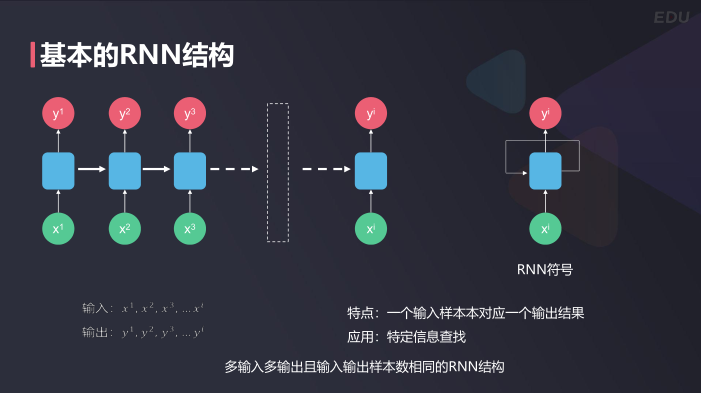
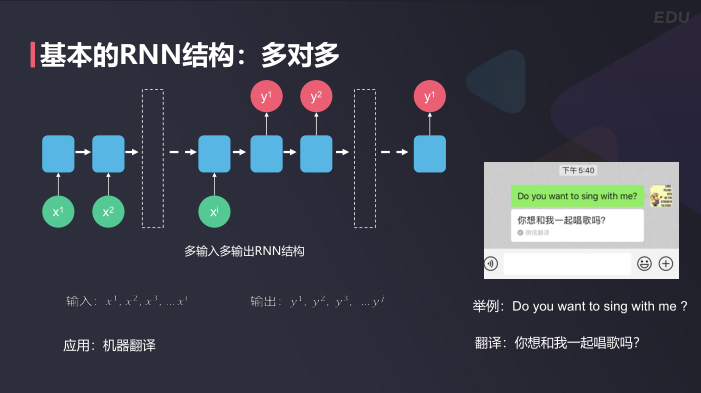
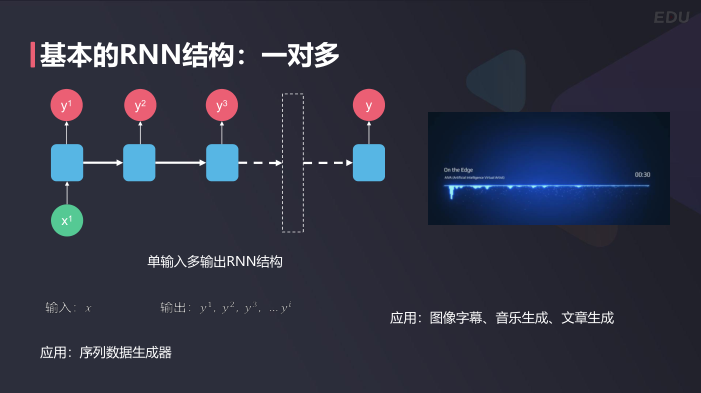
* **RNN处理字符串的方式：**
  1. **第一步词汇数值化：建立一个词汇-数值的字典，然后把输入词汇转化为数值数据（以下是两种字典创建方式，一般选择第一种字典创建方式**



* 1. **第二步转化为one-hot向量格式：不同字符之间不存在定量关系，因此需要通过one-hot向量格式的数据来表达字符信息，使数值之间没有量化关系。**



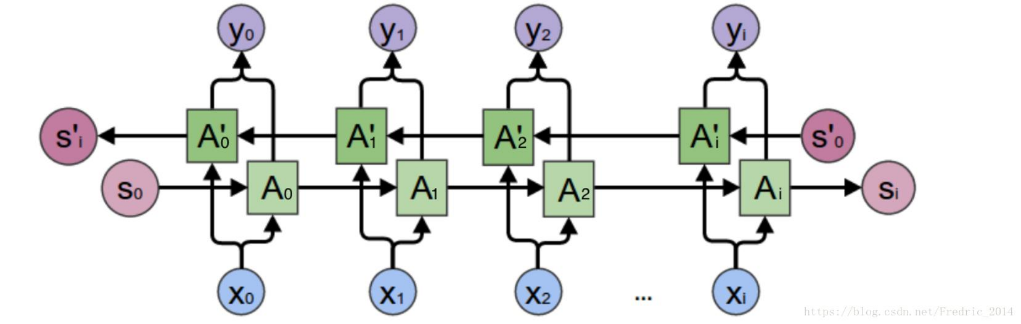
* **多种RNN结构：**
  1. **基本RNN结构：包括一对一基本RNN结构、多对一基本RNN结构、一对多基本RNN结构、多对多基本RNN结构**

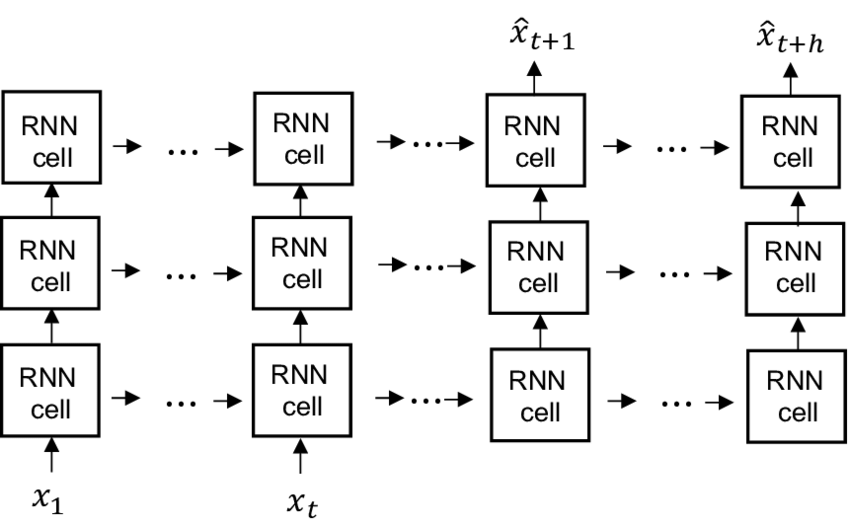
* 1. **长短期记忆网络（LSTM）：**
     1. **使用背景：基本RNN结构中的信息丢失问题——由于前部序列信息在传递到后部的同时，信息权重下降，导致部分重要信息可能会丢失；同时反向传播算法进行模型求解时，梯度也因此消失。从而导致基本RNN结构中会产生信息丢失问题。为解决RNN结构中的信息丢失问题，需要寻找一个一个保留重要信息决策权重的方法，即LSTM。**
     2. **LSTM的基本结构：**



* 1. **双向循环神经网络（BRNN）：**
     1. **使用背景：RNN和LSTM都只能依据之前时刻的时序信息来预测下一时刻的输出，但在有些问题中，当前时刻的输出不仅和之前的状态有关，还可能和未来的状态有关系。比如预测一句话中缺失的单词不仅需要根据前文来判断，还需要考虑它后面的内容，真正做到基于上下文判断。BRNN有两个RNN上下叠加在一起组成的，输出由这两个RNN的状态共同决定。**
     2. **BRNN的基本结构：增加一个反向传递的神经元，使后部序列信息也可用于前部信息的预测**

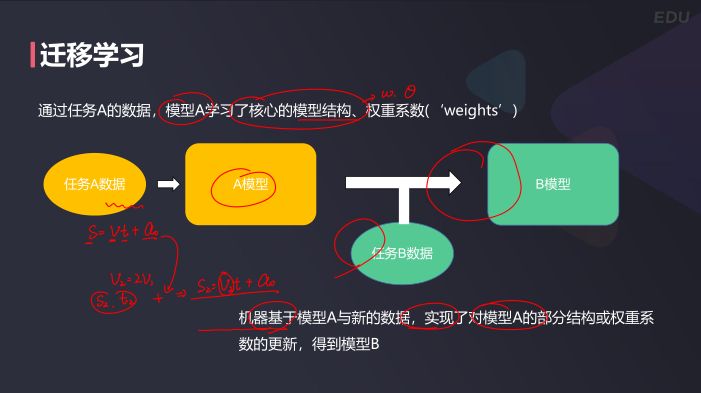
****

* 1. **深层循环神经网络（DRNN）：**
     1. **使用背景：对于更难提取的复杂信息，可以把单层RNN叠加起来或和MLP结构结合使用。**

****

**\*\*\*\*\*\*\*\*\*\*\*\* 混合模型 \*\*\*\*\*\*\*\*\*\*\*\***

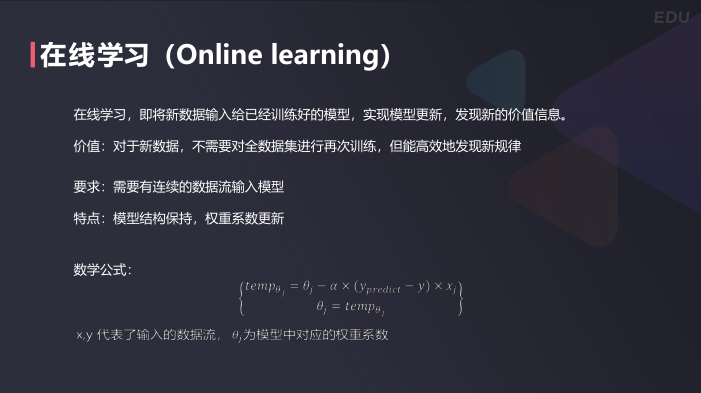
* **迁移学习（transfer learning）：迁移学习是一种机器学习方法，就是把为任务 A 开发的模型作为初始点，重新使用在任务 B 模型开发的过程中。**



* 1. **迁移学习的目的：运用已有的知识帮助发现新信息中的规律，并将其应用到解决不同但相似领域的问题。**
  2. **迁移学习的三种实现方法：**
     1. **特征提取：使用模型A，对原数据进行特征信息提取（通过会移除模型A的预测输出层）**
     2. **结构引用：直接使用模型A的结构，利用新数据对其重新重新/二次训练，实现权重系数更新**
     3. **部分训练：直接使用模型A的结构，利用新数据重新/二次训练，更新其指定的部分权重系数**
     4. **三种实现方法的使用场景：**



* 1. **迁移学习的价值：**
     1. **迁移学习对u样本数据集量需求相对较小**
     2. **待更新参数少，训练速度快**
     3. **通过会比很多全新的模型有更好的表现**
* **在线学习（Online learning）：在线学习，即将新数据输入给已经训练好的模型，实现模型更新，发现新的价值信息。**



* **半监督学习（Semi-Supervised Learning）：半监督学习是监督学习与无监督学习相结合的一种学习方法。半监督学习使用大量的未标记数据，以及同时使用标记数据，来进行模式识别工作。**
  1. **半监督学习的意义：在标记样本有限的情况下，尽可能识别出同类的共同特性，同时将大量无标签样本利用起来，发挥它们的价值。——寻找所有样本中的潜在规律，利用标签信息提高预测准确率**
  2. **半监督学习方法之一——伪标签学习（pseudo-labelling）：**
     1. **实现方法：用有标签数据训练一个分类器，然后用这个分类器对无标签数据进行分类，这样就会产生伪标签（pseudo label）或软标签（soft label），挑选你认为分类正确的无标签样本（预先设定一个挑选准则），把选出来的无标签样本用来训练分类器。**
     2. **局限性：①标签数据可能都是同一个类别，无法训练出有效的分类器；②可能无法寻找到合适的筛选准则**
* **混合模型**

