

南京邮电大学
Nanjing University of Posts and Telecommunications

数字图像与视频处理

第11章 图像识别





本章学习目标

- 掌握图像识别系统的框架结构，熟悉数据获取、预处理、特征提取以及分类器等各个模块的作用。
- 了解经验风险最小化和结构风险最小化的含义以及它们之间的区别，掌握支持向量机（SVM）分类方法。
- 掌握人工神经元模型，熟悉常见的人工神经网络，了解深度学习的概念。



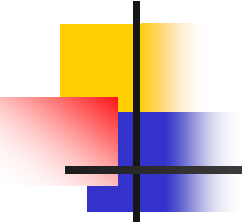
第11章 图像识别

- **11.1 图像识别概述**
- 11.2 统计学习方法
- 11.3 人工神经网络
- 11.4 基于LeNet网络的手写数字识别
- 11.5 MATLAB编程实例



11.1 图像识别概述

- 图像识别系统就是想让计算机也能够像人一样，识别出场景中感兴趣的目标。
- 设计一个图像识别系统，通常要涉及到如下的模块：
图像获取、数据预处理、特征提取、分类器设计、分类决策等。

- 
- 整个系统涉及到图像处理的三个层次——低层、中层和高层。一个传统的识别系统如图11-1所示：

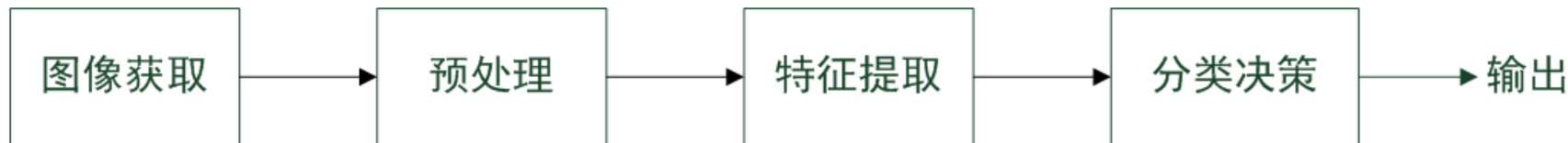


图11-1 传统的图像识别系统的基本构成



(1) 图像获取

- 图像获取是指通过光学摄像机、红外摄像机或激光、超声波、雷达等对现实世界进行传感，使计算机得到与现实世界相对应的二维或高维图像。
- 这些图像往往表示成数字形式，方便后续模块的处理。



(2) 预处理

- 预处理目的是去除噪声，加强有用信息，剔除干扰信号，并对输入测量仪器或其他因素所造成的退化现象复原。
- 涉及原理包括平滑、增强、复原、变换等技术。通过预处理后，为特征的正确、方便、和完整地获取提供可能。
- 图像预处理属于低层的操作。



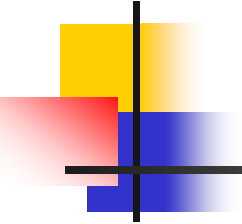
(3) 特征提取

- 为了有效地实现分类识别，就要对原始数据进行变换，得到最能反映分类的本质特征。
- 一般地，我们把原始数据所在的空间称为测量空间，把分类识别依赖进行的空间称为特征空间。通过变换，可把在维数较高的测量空间中表示的模式变为在维数较低的特征空间中表示的模式。
- 在特征空间中，一个模式通常也叫做一个样本，它往往表示为一个向量，即特征空间中的一个点。



(4) 分类决策

- 分类决策就是在特征空间中，利用分类器把待识别对象判决为某一个类别。
- 分类方法包括基于模板、基于统计理论、基于神经网络和基于聚类等多种。确定分类方法后，往往需要对这些方法中涉及到的参数进行设置。这个过程称为训练或者学习。
- 通过训练样本来训练分类器，使得根据这些参数来进行分类决策时，造成的错误识别率最小或引起的损失最小。

- 
- 从上述构成可以看出，传统的图像识别方法把特征提取和分类器设计分开，在应用时再将它们合在一起。
 - 比如如果输入是某种动物（如猫）的一系列训练图像，首先要对这些图像的特征进行提取，然后把表达出来的特征送到学习算法中进行训练得到分类器。



传统的识别方法缺点：

- 需要手工设计和提取特征，这需要大量的经验，需要对这个领域和数据特别了解，然后设计出来特征还需要大量的调试工作。
- 要有一个比较合适的分类器算法。只有特征和分类器是有效的，同时协调一致工作，才能使得系统识别达到最优。



基于深度学习的识别系统：

- 不手动设计特征，不挑选分类器
- 只需要将大量需要训练的图像以及这些图像的正负样本类型输入系统，系统自动完成特征提取和分类器的学习，然后将待识别图像输入系统，系统将直接输出识别结果。



图11.2 基于深度学习的图像识别系统的基本构成



第11章 图像识别

- 11.1 图像识别概述
- **11.2 统计学习方法**
- 11.3 人工神经网络
- 11.4 基于LeNet网络的手写数字识别
- 11.5 MATLAB编程实例



11.2.1 经验风险最小化

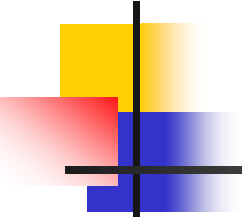
$$R(\theta) = \int L(y, f(x, \theta)) dF(x, y)$$

其中： $R(\theta)$ ： 期望风险

x ： 输入

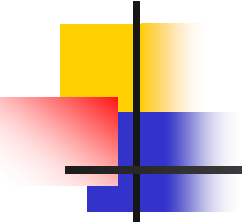
$y = f(x, \theta)$ ： 输出

$L(y, f(x, \theta))$ ： 损失函数

- 
- 机器学习的目的必须要求使得期望风险最小化，从而得到需要的目标函数。
 - 不难想象，可以利用给定的样本集上的平均损失最小化来代替无法求得的期望风险最小化。
 - 利用已知的经验数据（训练样本）来计算得到的误差，被称之为经验风险

$$R_{emp}(\theta) = \frac{1}{N} \sum_{i=1}^N L(y_i, f(x_i, \theta))$$

其中， $R_{emp}(\theta)$ 为经验风险。

- 
- 使用对参数求经验风险来逐渐逼近理想的期望风险的最小值，就是我们常说的经验风险最小化（**Empirical Risk Minimization, ERM**）原则。
 - 当样本容量足够大时，经验风险最小化能保证有很好的学习效果，在现实中被广泛采用。



11.2.2 结构风险最小化

- 结构风险最小化（**Structural Risk Minimization, SRM**）是为了防止过拟合而提出的策略。结构风险最小化等价于正则化。结构风险在经验风险的基础上加上表模型复杂度的正则化项：

$$R_{srn}(\theta) = \frac{1}{N} \sum_{i=1}^N L(y_i, f(x_i, \theta)) + \lambda J(f(x, \theta))$$

其中， $J(f)$ 为模型的复杂度。

11.2.3 支持向量机

- 支持向量机（Support Vector Machine, SVM）

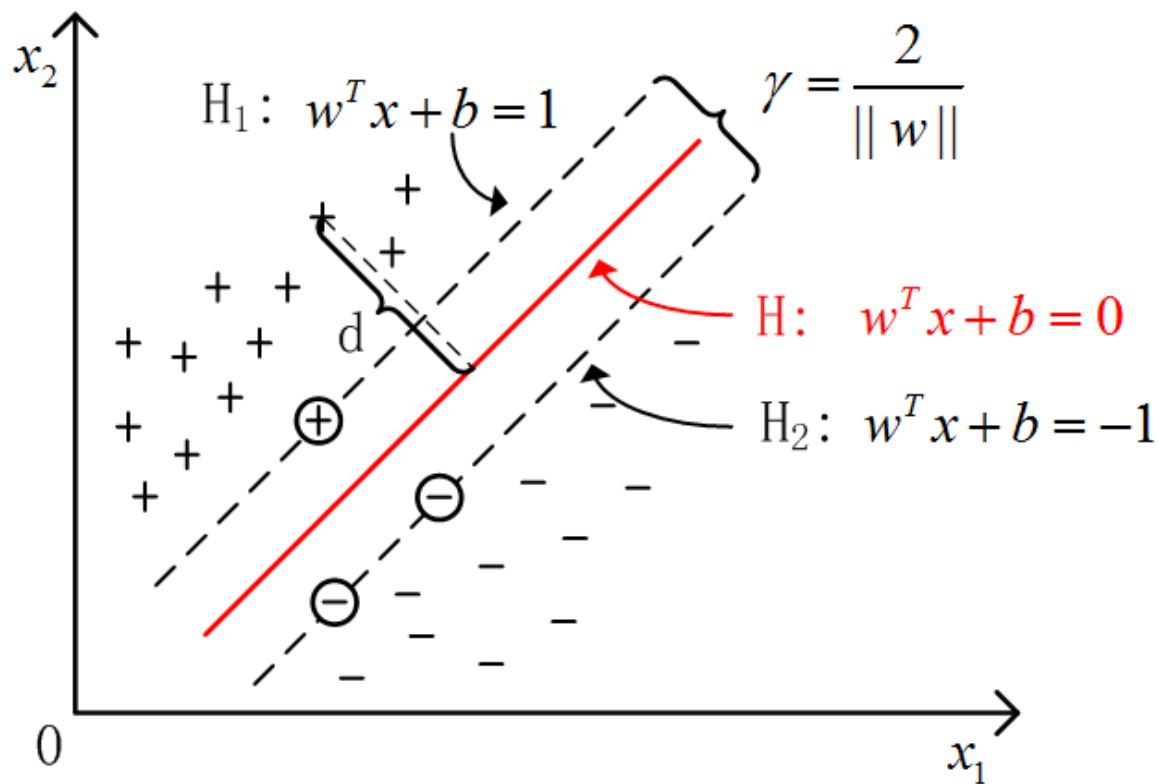
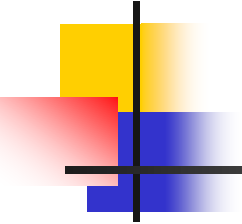
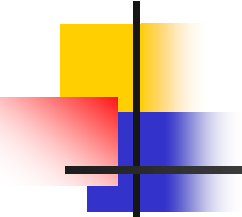


图11-3 最优分类示意图

- 
- 定义最优分类线 H ，可以使得分类间隔最远。
 - 分类间隔指的是图中 H_1 和 H_2 之间的距离。 H_1 和 H_2 分别是穿过正负样本离 H 最近的平行线。把二维的概念扩展到超平面上，最优分类线也就成了最优超平面。
 - 定义超平面的公式如下

$$f(x) = w^T x + b$$

- 
- 最优超平面的表示形式多种多样，通常用下面的表达式来表达最优超平面

$$|w^T x + b| = 0$$

- 支持向量（**support vector**）：假设 x 是距离超平面最近的一些点，也就是图中带有圈的点，这些点满足

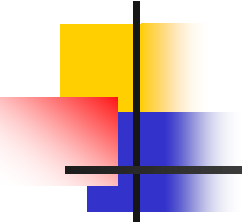
$$\begin{cases} (w^T x_i + b) = +1 & y_i = +1 \\ (w^T x_i + b) = -1 & y_i = -1 \end{cases}$$

- 
- 样本空间中任意一个点 x 到超平面 (w, b) 的距离 d 为

$$d = \frac{|w^T x + b|}{\|w\|}$$

- 定义 γ 为间隔 (margin)，其取值为最近距离的2倍

$$\gamma = \frac{2}{\|w\|}$$

- 
- 为了找到具有“最大间隔”（maximum margin）的划分超平面，也就是要找到约束参数 w 和 b ，使得 γ 最大，即

$$\max_{w,b} \frac{2}{\|w\|} \quad s.t. \quad y_i(w^T x_i + b) \geq 1 \quad i = 1, 2, \dots, m$$

- 最大化间隔，仅需要最大化 $\|w\|^{-1}$ ，等价最小化 $\|w\|^2$

$$\min_{w,b} \frac{1}{2} \|w\|^2 \quad s.t. \quad y_i(w^T x_i + b) \geq 1 \quad i = 1, 2, \dots, m$$

这就是SVM的基本型。

- SVM是一个线性分类器，但是它也可以把特征分类运用到非线性分类中。使用的方法是内核映射的方法。

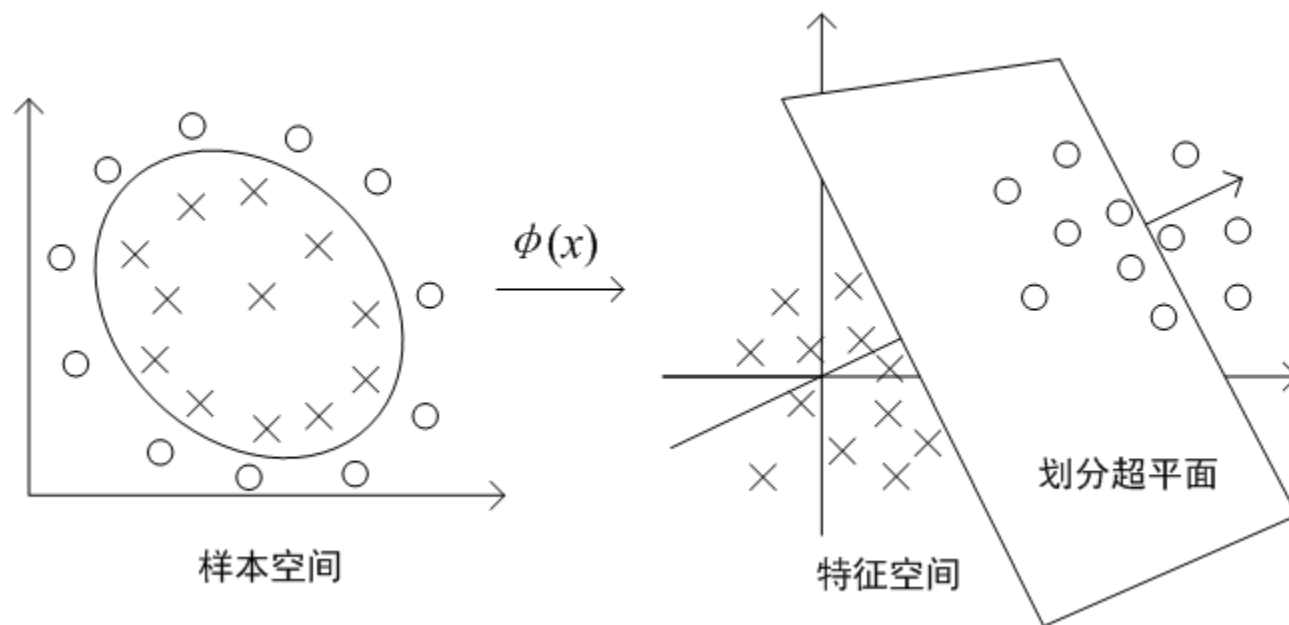
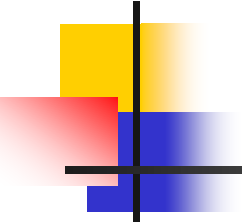


图11-4 从样本空间到特征空间的映射

- 
- 令 ϕ 是从原始样本空间 X 到特征空间 F 的映射, $\phi(x)$ 表示将 x 映射后的特征向量, 于是在特征空间进行划分超平面可以表示为

$$f(x) = w^T \phi(x) + b$$

- 此时优化问题为

$$\min_{w,b} \frac{1}{2} \|w\|^2 \quad s.t. \quad y_i (w^T \phi(x_i) + b) \geq 1 \quad i = 1, 2, \dots, m$$



第11章 图像识别

- 11.1 图像识别概述
- 11.2 统计学习方法
- **11.3 人工神经网络**
- 11.4 基于LeNet网络的手写数字识别
- 11.5 MATLAB编程实例



11.3 人工神经网络

- 人工神经网络（**Artificial Neural Network, ANN**）高度非线性网络，可用来描述认知、决策及控制等智能行为，在目标检测、物体分类以及识别等领域取得了成功。
- 它的研究和人脑结构的认识和研究有着密切关系。



11.3.1 人工神经元模型

- 神经元是大脑处理信息的基本单元。
- 它主要由细胞体、树突、轴突和突触（Synapse，又称“神经键”）组成。
- 一个神经元有许多输入端（当然也有较少的输入，完成中继放大的作用），即突触，每个突触的大小可以是不同的，也就是它们由接受输入脉冲到刺激本神经元的细胞膜的强度是不一样的。



人工神经元有如下三个基本要素：

- 连接强度。用来与其他神经元的连接，模拟生物神经元的突触。
- 求和单元。计算当前神经元的所有输入信号的加权和。
- 激励函数（传递函数）。将加权信号映射为输出信号。

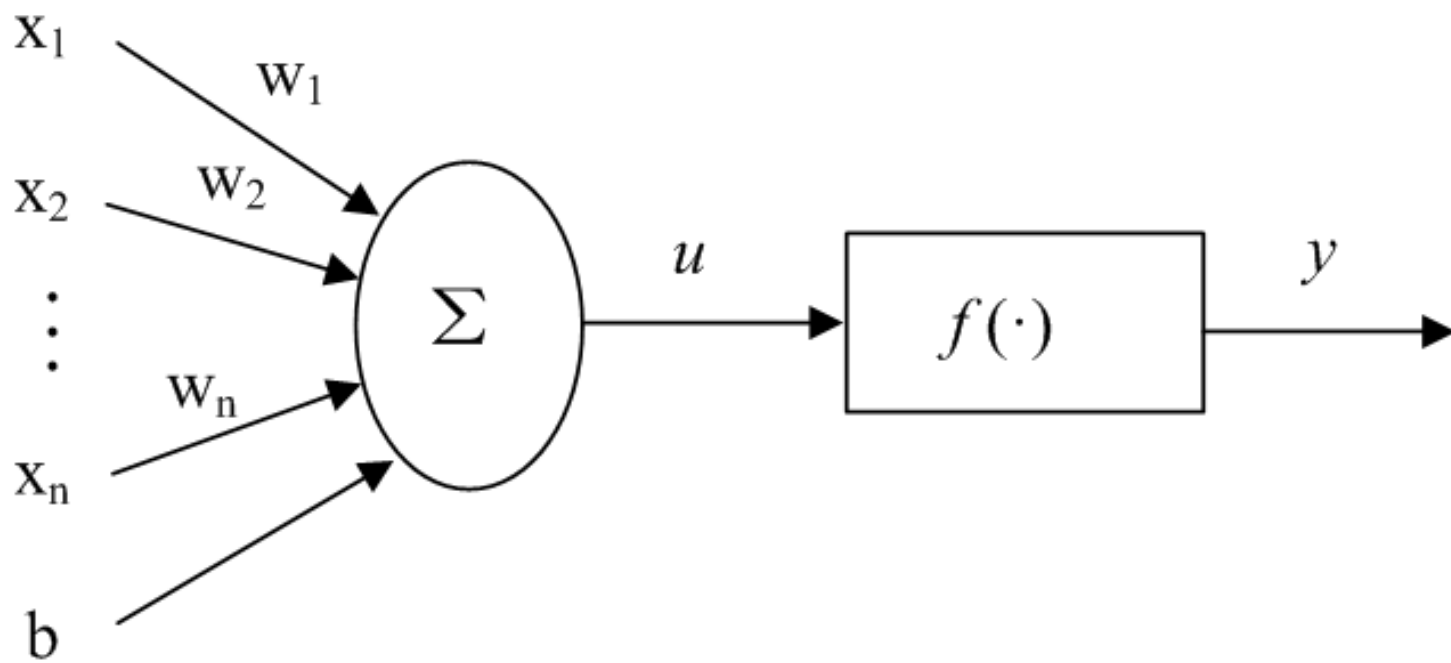


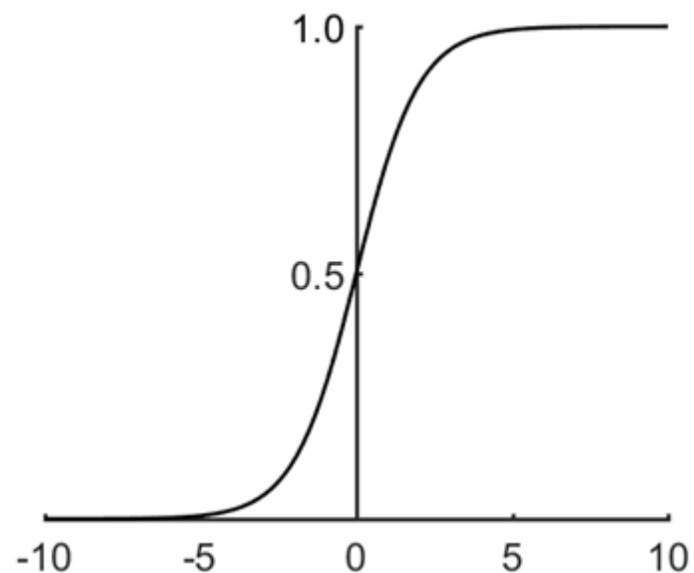
图11-5 人工神经元结构模型



常用的激活函数

□ Sigmoid函数

$$f(x) = \frac{1}{1 + e^{-x}}$$

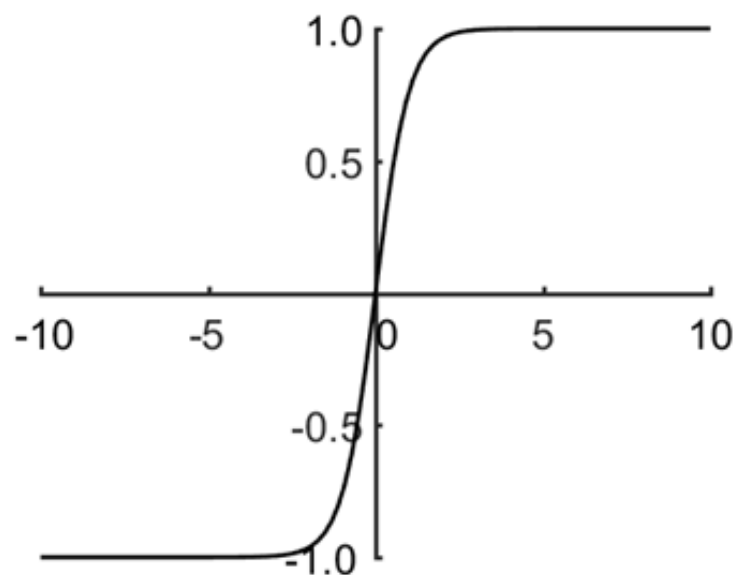




常用的激活函数

□ Tanh函数

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

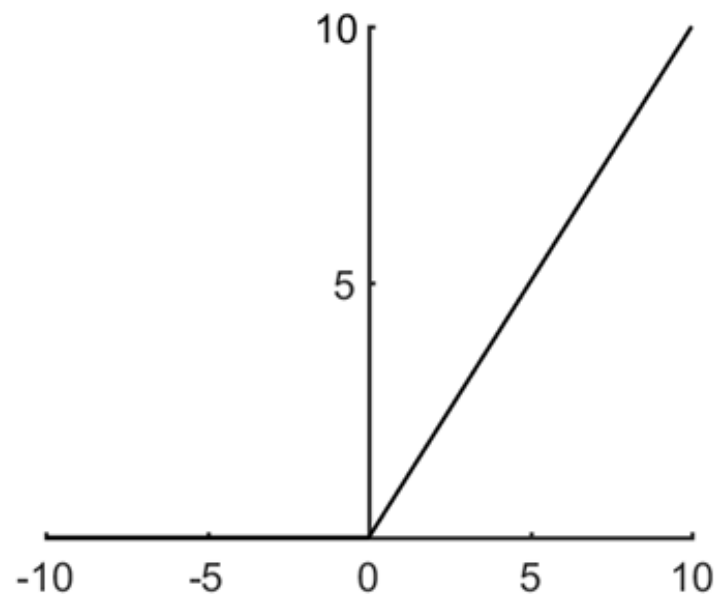




常用的激活函数

□ ReLU函数

$$f(x) = \max(0, x)$$





11.3.2 前馈神经网络

前馈神经网络每个神经元接受前一级输入，并输出到下一级，层间无反馈。

- (1) 感知器

最简单的前馈网络在1958年由Rosenblatt等人提出的感知器（Perceptron，也称为感知机）模型。它是一种两层神经网络，即输入层和输出层。

- 三层以及三层以上前馈神经网络通常又被称为多层感知器。

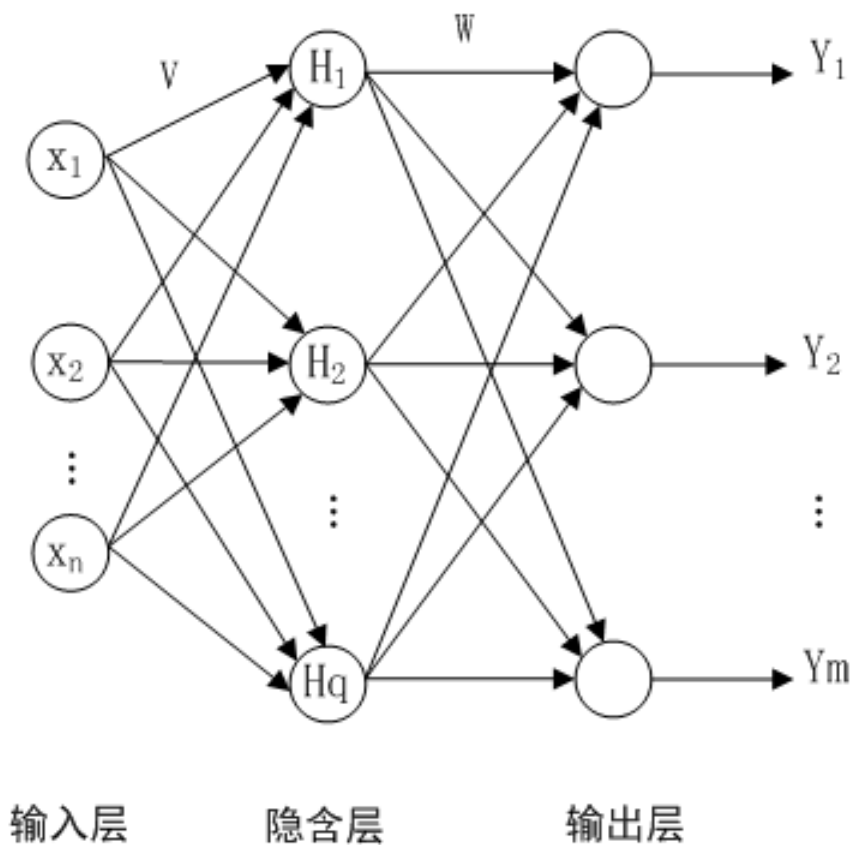


图11-9 含一个隐含层的前馈神经网络



■ (2) RBF网络

这种网络只有一个隐层，隐层单元采用径向基函数作为其激活函数，输入层到隐层之间的权值固定为1，输出节点为线性求和单元，隐层到输出节点之间的权值可调，因此输出为隐层的加权求和。

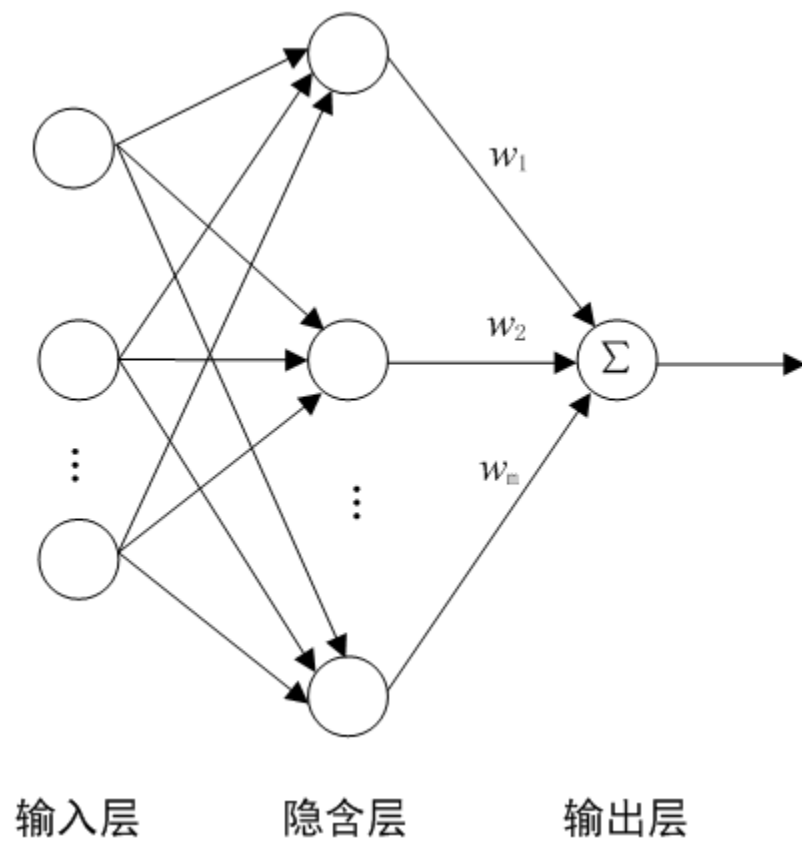
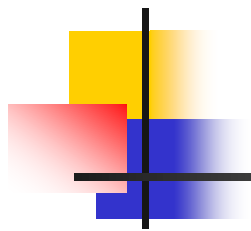


图11-10 RBF网络结构



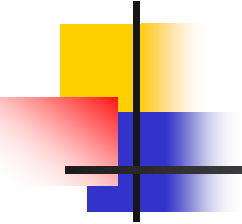
11.3.3 Hopfield网络

- 前馈网络不考虑输出与输入之间在时间上的滞后影响，其输出与输入之间仅仅是一种映射关系。
- Hopfield网络采用反馈连接，所有神经元之间相互连接，考虑输出与输入在时间上的传输延迟，所表示的是一个动态过程，需要用差分或微分方程来描述。
- Hopfield网络权值对称，通常也没有自反馈。



11.3.4 卷积神经网络

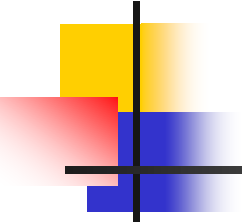
- 深度学习是机器学习研究领域的一个新的分支，是一类复杂的机器学习算法。
- 其研究的目的在于建立、模拟人脑的神经网络，并模仿人脑的机制来解释如图像、声音和文本之类的数据。
- 卷积神经网络（**Convolutional Neural Networks, CNN**）是深度学习中应用较为广泛的一种模型。

- 
- 深度学习之所以被称为“深度”，是相对SVM、提升算法(boosting)、最大熵方法、以及只含单隐层的多层感知器等“浅层学习”方法而言。
 - 其实质是通过搭建具有多个隐层的学习模型，给其输入海量的训练数据，使其从训练数据中学习获得有用的特征，从而最终提升分类或预测的准确性。
 - “深度模型”是手段，“特征学习”是目的。



与传统的浅层学习相比，深度学习的不同在于

- 模型有多个隐层，一般比较深，通常有5层、6层，甚至成百上千层；
- 模型能够从训练数据中自主提取特征。浅层学习依靠人工经验抽取样本特征，模型只用来负责分类和预测。深度学习自动地学习得到层次化的特征表示。

- 
- 1958年，Hubel和Wiesel对猫视觉皮层电生理的研究激发了人们对于人类神经系统的思考，Fukushima受此启发提出了卷积神经网络的模型。
 - 与传统神经网络不同的是，卷积神经网络在卷积阶段使用了局部感受野和权值共享策略来减小网络参数。
 - 此外，卷积网络在卷积层后面引入下采样层，可以扩大感受野的同时降低网络的参数，实现平移不变性。

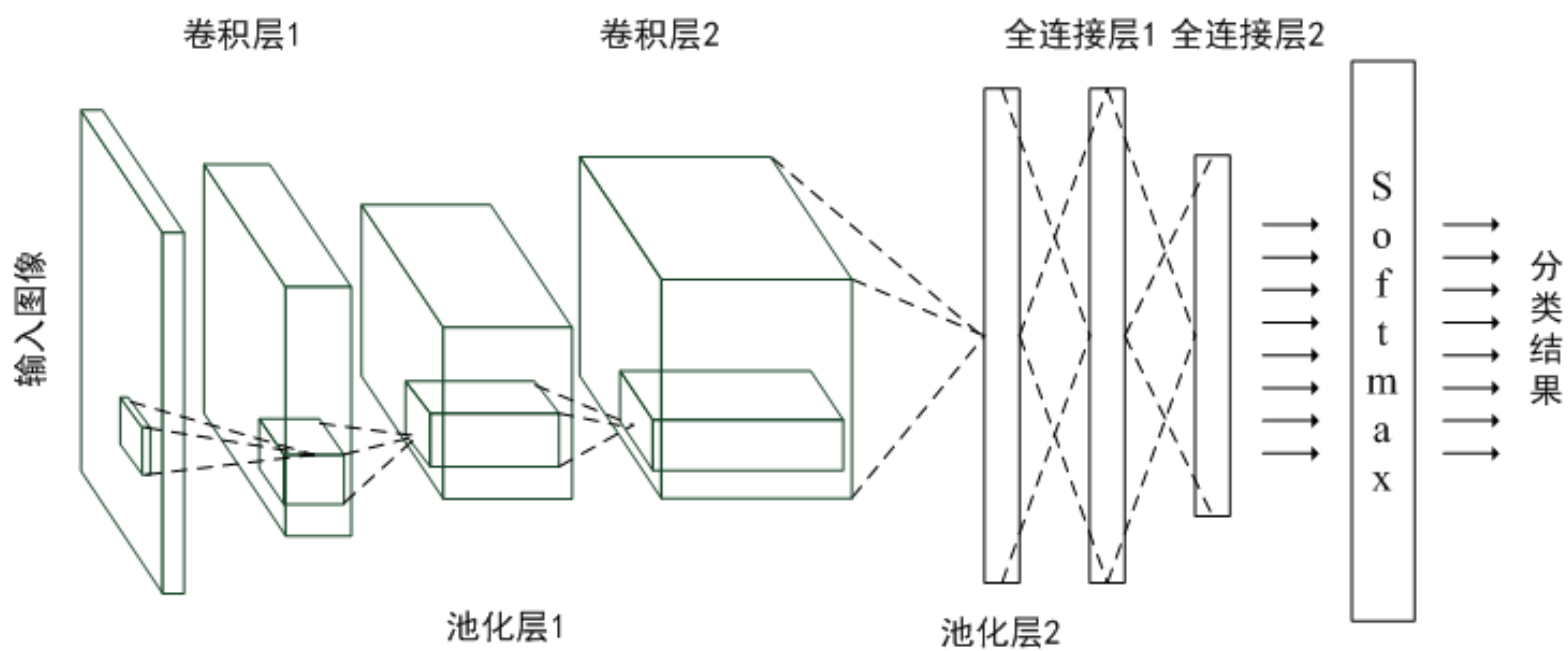


图11-11 CNN网络基本架构



各个网络部分的结构的功能如下：

- (1) 输入层：输入层是即送入网络的输入数据，在图像识别中，就是一幅图像数据矩阵。
- (2) 卷积层：卷积神经网络的卷积层也称为特征提取层，它往往用多个不同的卷积核（权重参数不同）来卷积，可以认为不同的卷积核从输入数据中提取的特征不相同。

x ₁₁	x ₁₂	x ₁₃	x ₁₄	x ₁₅
x ₂₁	x ₂₂	x ₂₃	x ₂₄	x ₂₅
x ₃₁	x ₃₂	x ₃₃	x ₃₄	x ₃₅
x ₄₁	x ₄₂	x ₄₃	x ₄₄	x ₄₅
x ₅₁	x ₅₂	x ₅₃	x ₅₄	x ₅₅

(a) 输入特征图

*

w ₁₁	w ₁₂	w ₁₃
w ₂₁	w ₂₂	w ₂₃
w ₃₁	w ₃₂	w ₃₃

(b) 卷积核

=

y ₁₁	y ₁₂	y ₁₃
y ₂₁	y ₂₂	y ₂₃
y ₃₁	y ₃₂	y ₃₃

(c) 输出特征图

$$\text{其中, } y_{ij} = \sum_{m=1}^3 \sum_{n=1}^3 (x_{i+m-1, j+n-1} \times w_{mn})$$

图11-12 二维卷积示意图

- (3) 采样层：也称为下采样层，或者pooling层（池化层）。

主要的目的就是在保留有用信息的基础上减少数据的处理量，加快网络的训练速度。

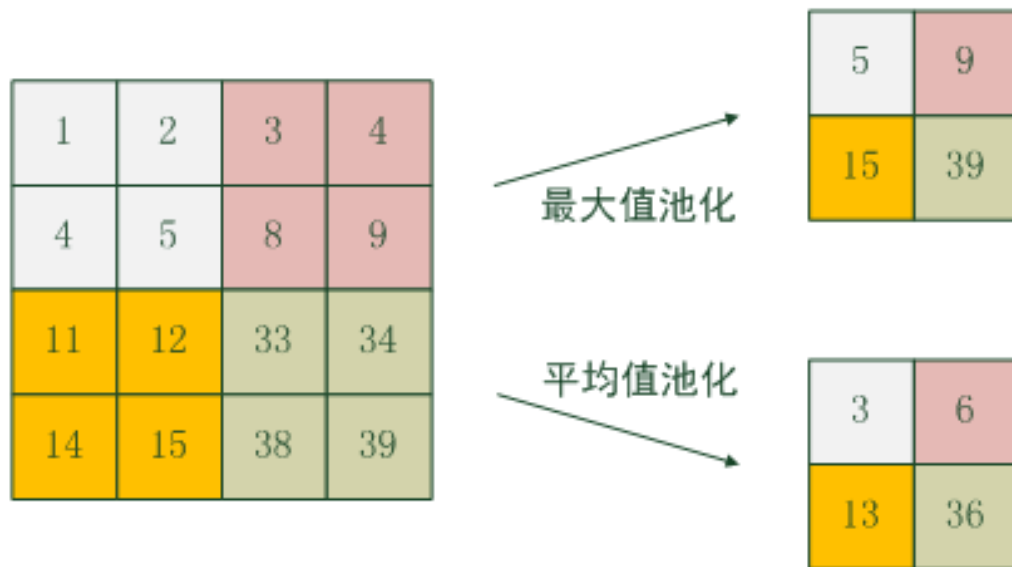
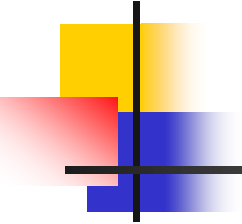


图11-13 最大值池化和平均值池化示意

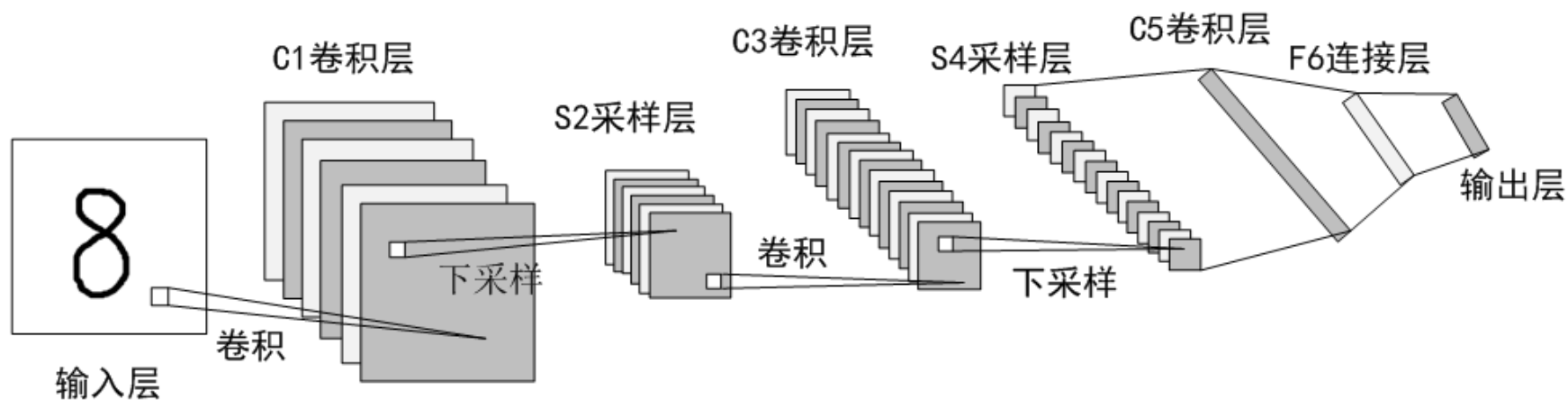
- 
- (4) 全连接层：经过几轮卷积层和池化层之后，可以认为图像中信息已经转变为高度抽象特征。在特征提取完成后，需要全连接层完成分类任务。
 - (5) 输出层：输出层的神经元节点需要根据具体任务而进行具体设定。对图像分类任务，网络输出层为一个分类器，借助softmax可以得到当前样例属于不同类别概率分布。

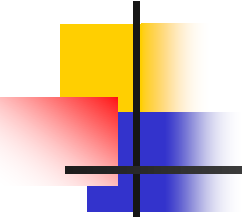


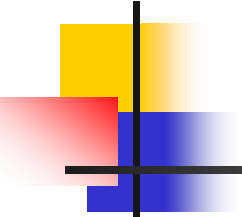
第11章 图像识别

- 11.1 图像识别概述
- 11.2 统计学习方法
- 11.3 人工神经网络
- **11.4 基于LeNet网络的手写数字识别**
- 11.5 MATLAB编程实例

11.4 基于LeNet网络的手写数字识别



- 
- 第一层输入层是 32×32 大小的图像。
 - 第二层C1层为卷积层，包括6个特征图，卷积核大小为 5×5 。
 - 第三层S2为下采样层，有6个 14×14 大小的特征图，每个特征图的每个神经元都和C1层对应的特征图的 2×2 区域相连接。
 - 第四层C3层为卷积层，有16个特征图，采用 5×5 卷积核。

- 
- 第五层S4为下采样层。有16个 5×5 的特征图组成。每个神经元和C3中特征图的 2×2 区域相连。
 - 第六层C5是卷积层，使用 5×5 的卷积核。每个特征图有1个神经元。每个神经元和S4层的全部16个特征图的 5×5 的区域全连接。
 - 第七层F6为全连接层，有84个特征图，每个特征图只有一个神经元和C5层全连接。
 - 第八层为输出层，也是全连接层，共有10个节点，分别代表数字0~9。