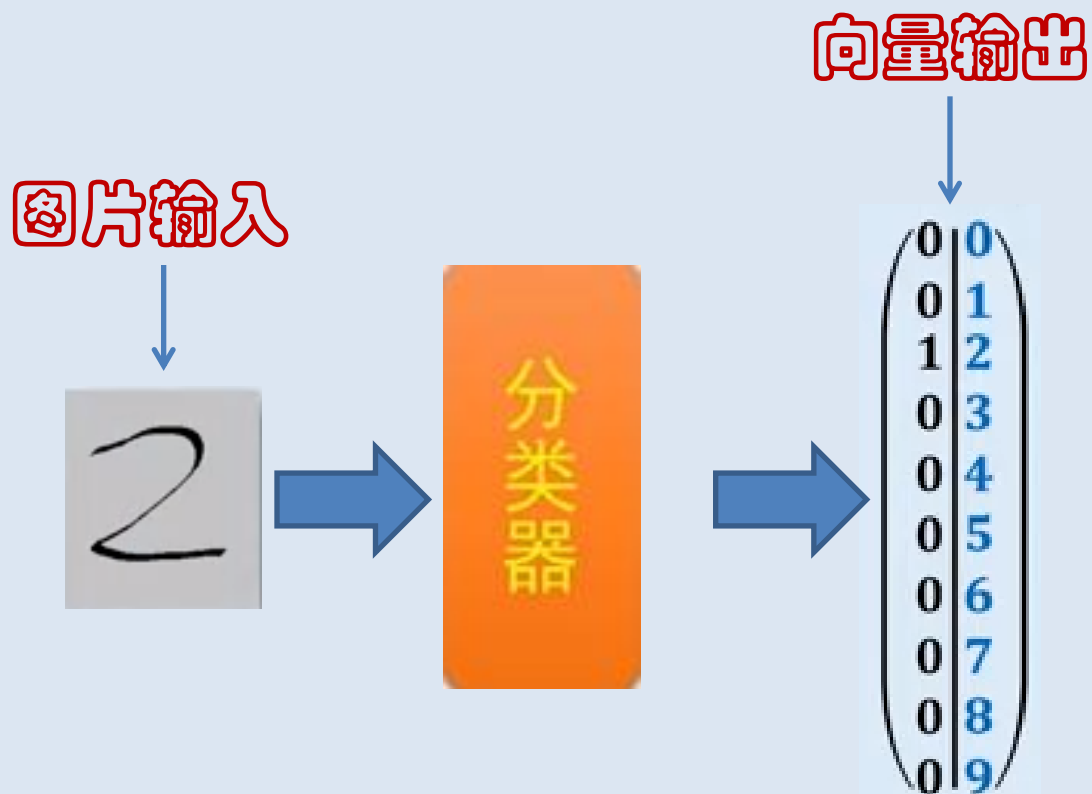


手写样式千差万别
笔的粗细、软硬不同
书写位置、朝向不同
背景的种种干扰.....

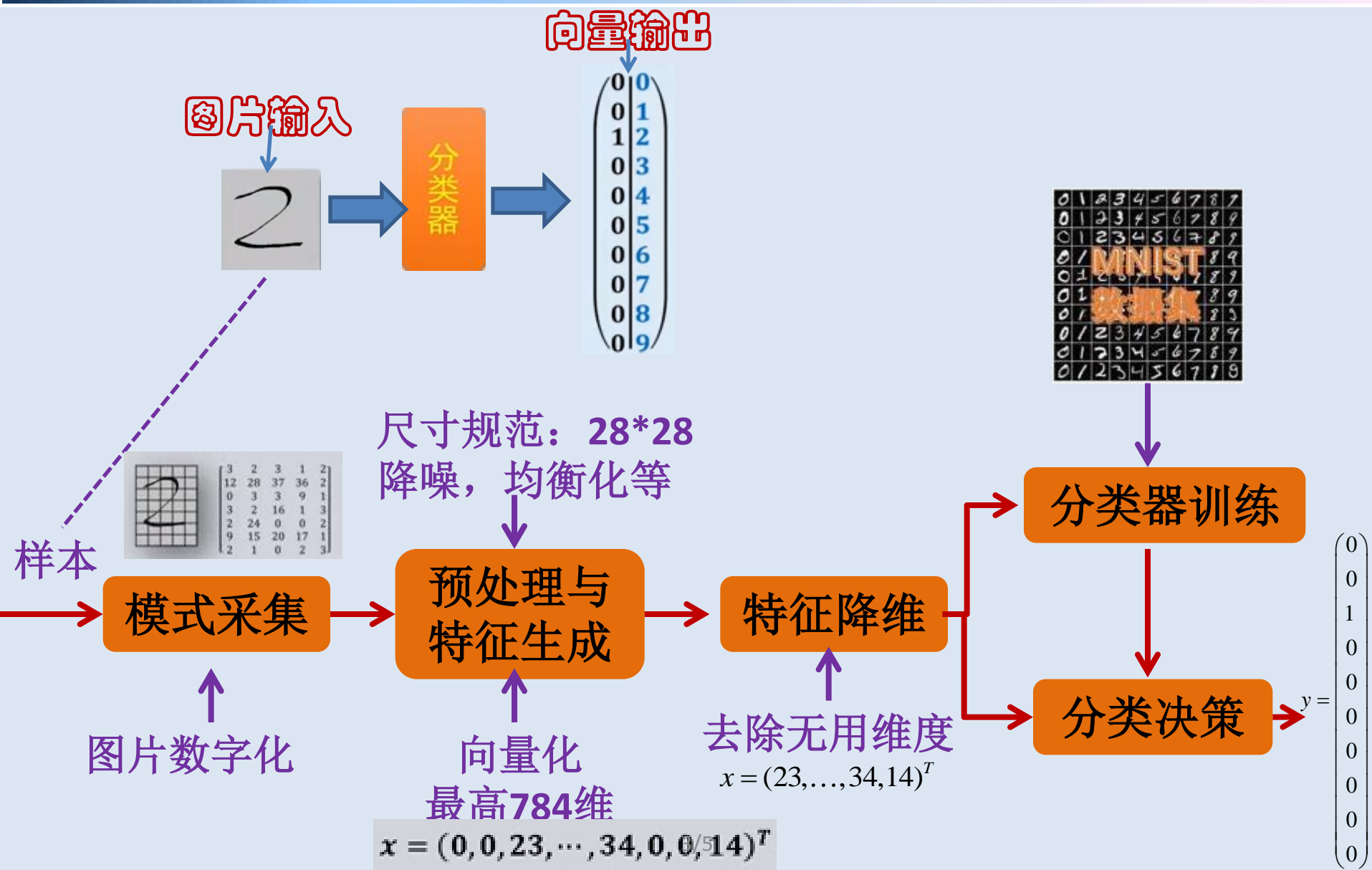


手写数字识别





手写数字识别





mnist数据集



mnist数据集：来自美国国家标准与技术研究所（NIST）

下载网址：<http://yann.lecun.com/exdb/mnist/>

- 训练集：由来自 250 个不同人手写的数字构成，其中 50% 是高中学生，50% 来自人口普查局的工作人员，共60000个训练样本

模式识别和机器学习领域的典范数据集

- 测试集：同样比例的手写数字，共10000个测试样本

日新学堂资料版块已上传 matlab矩阵mat格式的mnist数据集

共四个文件：训练集 train_images.mat、测试集test_images.mat以及它们的标签

✓ 图片数据为三维矩阵： $28 \times 28 \times 60000$ （ $28 \times 28 \times 10000$ ）

每一张 28×28 图片均为单通道（灰度），包含784个数据点，每个数据点代表当前位置的灰度像素值，范围0-255

✓ 标签数据为行向量： 1×60000 （ 1×10000 ）

不同于标准mnist数据集标签，已将原始的十维向量转换为0-9的数字标签如（原始0100000000转换为数字1）



一 基于贝叶斯分类器的手写数字识别

五一假期前提交

实现一个基于贝叶斯分类器的手写数字识别方法 (`fitcnb`)

提交一个电子文档（需写上姓名和学号），包括如下部分：

(1) 实现代码

(2) 用表格或者图像的方式展示训练样本数量和分类准确率的关系。看看增加训练样本数量，分类准确率是否提高？是否可以无限提高？

训练样本数量： 2000、3000、4000、5000、6000、7000
(至少上述5种情况)

测试样本数量： 200 (300或者其它的)，在提交的文档中需要标明测试样本数量

不要抄袭！

若判定为抄袭，抄袭者和被抄袭者的本次作业成绩均记0分！