目 录

第 1 章

Introduction

The field of of pattern recognition is concerned with the automatic discovery of regularities in data through the use of computer algorithms and with the use of the regularities to take actions such as classifying the data into different categories.
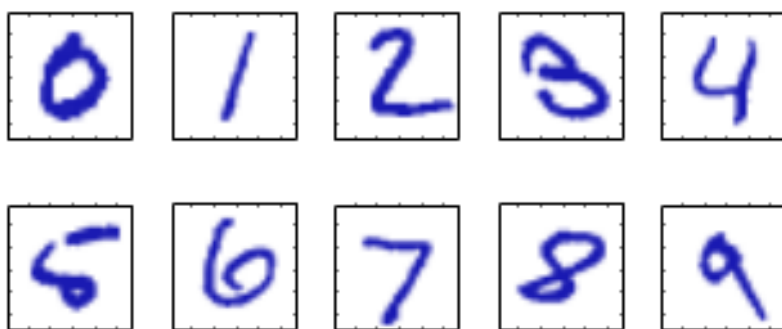
1.1 Handwritten digits recognition

图 1.1: 美国邮件上的手写数字样本

1.1.1 empirical approach

hand-crafted rules or heuristics for distinguishing the digits based on the shapes of the strokes. 使用手工编写的规则或者是启发 来通过笔画的形状鉴别阿拉伯数字。

Proliferation of rules and poor results.

1.1.2 machine learning approach

$\diamond$ N 位的训练集(tanning set) $\{x_1, \ldots, x_N\}$

$\diamond$ 目标向量(target vector)来表示每训练集元素对应的类别 (0-9)

$\diamond$ 训练过程(tanning phase)确定了函数 $\mathbf{y}(x)$

1.2 Example:Polynomial Curve Fitting

$$y(x, \mathbf{w}) = w_0 + w_1 + w_2 x^2 + \ldots + w_M x^M = \sum_{j=0}^{M} w_j x^j \tag{1.1}$$

$$E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^{N} y\{(x_n, \mathbf{w}) - t_n\}^2 \tag{1.2}$$

Definition:RMS

$$E_{RMS} = \sqrt{2E(\mathbf{w}^*)/N} \tag{1.3}$$

$$\tilde{E}(\mathbf{w}) = \frac{1}{2}\{y(x_n, \mathbf{w}) - t_n\}^2 + \frac{\lambda}{2}\|\mathbf{w}\|^2 \tag{1.4}$$

1.3 Probablility Theory

The rules of Probability

sum rule

$$p(X) = \sum_Y p(X, Y) \tag{1.5}$$

product rule

$$p(X, Y) = p(Y|X)p(X) \tag{1.6}$$

Bayes'theorem

$$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)} \tag{1.7}$$

贝叶斯推论

$$P(X) = \sum_Y p(X|Y)p(Y) \tag{1.8}$$

Note: 可以看成原始兩邊對 Y 累加，然后左邊 $P(Y|X)$ 項累加和等于 1 約去

1.3.1 Probability densities

x 上 (a,b) 区间内的概率密度

$$p(x \in (a,b)) = \int_a^b p(x)dx \tag{1.9}$$

概率密度(probability density)$p(x)$ **必须满足的两个条件**

1.

$$p(x) \geqslant 0 \tag{1.10}$$

2.

$$\int_{-\infty}^{\infty} p(x)dx = 1 \tag{1.11}$$

$$p_y(y) = p_x(x)|\frac{dx}{dy}| \tag{1.12}$$
$$= p_x(g(y))|g'(y)|$$

⚗ *Note:* 額。。。。暂時没看懂

在 $(-\infty, z)$ 区间内的**概率分布函数**(cumulative distribution function)定义如下：

$$P(z) = \int_{-\infty}^z p(x)dx \tag{1.13}$$

⚗ *Note:* 满足了 $P'(x) = p(x)$

当有多个连续的自变量 $x_1, \ldots, x_D$ 时，将其整体表示为 $\mathbf{x}$，于是可以定义一个联合的概率密度函数 $p(\mathbf{x}) = p(x_1, \ldots, x_D)$ 于是有了以下两个条件：

1.

$$p(\mathbf{x}) \geqslant 0 \tag{1.14}$$

2.

$$\int_{-\infty}^{\infty} p(\mathbf{x})d\mathbf{x} = 1 \tag{1.15}$$

如果 x 是一个离散的变量，$p(x)$ 就被叫做**概率质量函数**(probability mass function)

$$p(x) = \int p(x,y)dy \tag{1.16}$$

$$p(x,y) = p(y|x)p(x) \tag{1.17}$$

1.3.2 Expectations and covariances

在概率分布 $p(x)$ 下的某些函数 $f(x)$ 的平均值叫做 $f(x)$ 的期望(expectation)，标记为 $\mathbb{E}[f]$
离散分布下期望定义为：

离散分布期望定义

$$\mathbb{E}[f] = \sum_x p(x)f(x) \tag{1.18}$$

连续分布下期望定义为：

连续分布期望定义

$$\mathbb{E}[f] = \int p(x)f(x)dx \tag{1.19}$$