PRML笔记

2017年12月12日 星期二 下午1:59

1 Introduction

example: recognizing handwritten digits 手写识别

training set:

a large set of N digits $\{x_1,...,x_N\}$

To tune the parameters of an adaptive model 用来调节模型的参数

target vector t/y:

function y(x):

determined during the *training/learning* phase,

Preprocessed

transform them into some new space of variables

speed up computation

1、*supervised learning*有监督学习 target vectors

1. *classification* 分类

   assign each input vector to one of a finite number of discrete categories,
2. *regression*. 回归

   output consists of one or more continuous variables

2、*unsupervised learning*无监督学习 without target vectors

1. *clustering* 聚类

   similar examples within the data
2. *density estimation* 密度估计

   determine the distribution of data within the input space 决定输入空间中数据的分布
3. *Visualization* 可视化

   high-dimensional space down to two or three dimensions

3、*reinforcement learning* 反馈学习

finding suitable actions to take in a given situation in order to maximize a reward.

在给定的条件下，找到合适的动作，使回报达到最大值

feature:

trade-off between *exploration* and *exploitation*

- *exploration*：探索
  - system tries out new kinds of actions to see how effective they are,
- *exploitation*：利用

- system makes use of actions that are known to yield a high reward

1.1 Example: Polynomial Curve Fitting 多项式曲线拟合

given a training set

$\mathbf{x} \equiv (x_1, \ldots, x_N)^{\mathrm{T}}$

observations of the values 观测值

$\mathbf{t} \equiv (t_1, \ldots, t_N)^{\mathrm{T}}$

fit the data using a polynomial function of the form 多项式函数来拟合数据

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

y(x, w) is a linear function of the coefficients w. 线性模型

error function 误差函数

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

choosing the order M

model comparison or *model selection*.

root-mean-square (RMS) error

$$E_{\mathrm{RMS}} = \sqrt{2E(\mathbf{w}^{\star})/N}$$

over-fitting problem 控制过拟合问题

Regularization 正则化

adding a penalty term to the error function

给误差函数增加一个惩罚项，

调节参数λ

$$\widetilde{E}(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^{N} \{y(x_n, \mathbf{w}) - t_n\}^2 + \frac{\lambda}{2} \|\mathbf{w}\|^2$$

1.2 Probability Theory 概率论

加法规则

乘法规则

贝叶斯定理

先验概率

后验概率

后验概率
概率密度 probability density

$$p(x \in (a,b)) = \int_a^b p(x)\ \mathrm{d}x$$

累积分布函数 cumulative distribution function

$$P(z) = \int_{-\infty}^{z} p(x)\ \mathrm{d}x$$

两者关系：

$P'(x) = p(x)$；

推出新的加法规则和乘法规则

$$p(x) = \int p(x,y)\ \mathrm{d}y$$

$$p(x,y) = p(y \mid x)p(x)$$

期望

离散

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

连续

$$\mathbb{E}[f] = \int p(x)f(x)\ \mathrm{d}x$$

方差

$$\mathrm{var}[f] = \mathbb{E}[(f(x) - \mathbb{E}[f(x)])^2]$$

1.2.3 Bayesian probabilities 贝叶斯定理

先验概率p(w) + 条件概率p(D | w)

$$p(\mathbf{w}|\mathcal{D}) = \frac{p(\mathcal{D}|\mathbf{w})p(\mathbf{w})}{p(\mathcal{D})}$$

posterior ∝ likelihood × prior

频率学家

最大似然(maximum likelihood)估计

自助法(bootstrap)

马尔科夫链蒙特卡罗 应用于小规模问题

马尔科夫链蒙特卡罗 应用于小规模问题

1.2.4 The Gaussian distribution 高斯分布

$$\mathcal{N}\left(x|\mu,\sigma^2\right)=\frac{1}{(2\pi\sigma^2)^{1/2}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

μ均值(mean)，

σ^2方差(variance)

方差的倒数，记作精度 (precision)

$$\beta=\frac{1}{\sigma^2}$$