Introduction

2017-12-06 14:23

An example

- Hand writing recognition: machine learning is a kind of method of
- Curve Fitting:
  - $E(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{y(x_n, \mathbf{w}) - t_n\}^2$
  - Root-mean-square error (RMS-Error)
    - $E_{\mathrm{RMS}} = \sqrt{2E(\mathbf{w}^\star)/N}$
- To prevent over fitting
  - More data
  - By regularization（正则化）
    - $\widetilde{E}(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{y(x_n, \mathbf{w}) - t_n\}^2 + \frac{\lambda}{2}\|\mathbf{w}\|^2$
- Suggest: taking the available data and partitioning it into a training set, used to determine the coefficients w, and a separate validation set, also called a hold-out set, used to optimize the model complexity (either M or λ). However it is a kind of wasting training data
- 这一部分作为导入很贴切，基本说明了是在用什么思路解决什么问题、解决问题中可能遇到的瓶颈以及常用的解决策略

Probability Theory

0.基本概念

- Sum rule of probability (加法法则 边缘概率)
  - $p(X = x_i) = \sum_{j=1}^{L} p(X = x_i, Y = y_j) \quad , = \frac{c_i}{N}.$

$\because$

$\therefore$

- conditional probability (条件概率)
  - $p(Y = y_j | X = x_i) = \frac{n_{ij}}{c_i}.$
- Joint probability (联合概率)
  - $$\begin{aligned} p(X = x_i, Y = y_j) &= \frac{n_{ij}}{N} = \frac{n_{ij}}{c_i} \cdot \frac{c_i}{N} \\ &= p(Y = y_j | X = x_i) p(X = x_i) \end{aligned}$$

p(X,Y) = p(X|Y)p(Y)=p(Y|X)p(X)=p(Y,X)

p(Y|X) = p(X|Y)p(Y)/p(X)

- Productive rule (乘法法则)
  - $p(X, Y) = p(Y|X)p(X).$
- Bayes' theorem
  - $p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)}$
- Prior probability (先验概率)：在随机变量的值确定前获得的概率。如从蓝色盒子中抽取的概率是40%
- Posterior probability (后验概率)：在随机变量的值确定后推算出的概率。如已知抽取到的是苹果，推算出之前是从蓝盒子中抽取的概率。

1.概率密度（当随机变量的取值是连续型时）

- The probability that x will lie in an interval (a, b) is then given by
  - $p(x \in (a, b)) = \int_a^b p(x)\,\mathrm{d}x.$
- 原本要求p(y)需要用p_y，在知道随机变量x和y之间的关系后，能将原问题转化成用p_x来求
  - Under a nonlinear change of variable, a probability density transforms differently from a simple function, due to the Jacobian factor. For instance, if we consider a change of variables $x = g(y)$, then a function $f(x)$ becomes $\widetilde{f}(y) = f(g(y))$. Now consider a probability density $p_x(x)$ that corresponds to a density $p_y(y)$ with respect to the new variable y, where the suffices denote the fact that $p_x(x)$ and $p_y(y)$ are different densities. Observations falling in the range $(x, x + \delta x)$ will, for small values of δx, be transformed into the range $(y, y + \delta y)$ where $p_x(x)\delta x \simeq p_y(y)\delta y$, and hence

？ 这里的约等于关系是怎么来的

$\left| \frac{\mathrm{d}x}{} \right|$

- are different densities. Observations falling in the range $(x, x+\delta x)$ will, for small values of δx, be transformed into the range $(y, y+\delta y)$ where $p_x(x)\delta x \simeq p_y(y)\delta y$, and hence

  $$\begin{aligned} p_y(y) &= p_x(x)\left|\frac{\mathrm{d}x}{\mathrm{d}y}\right| \\ &= p_x(g(y))\,|g'(y)|. \end{aligned} \tag{1.27}$$

  ？ 这里的约等于关系是怎么来的

- 如果随机变量的取值是离散的，那么将p(X)称为概率质量函数probability mass function
- 对于连续型变量，其概率的加法公式和乘法公式为
  - $$\begin{aligned} p(x) &= \int p(x,y)\,\mathrm{d}y \\ p(x,y) &= p(y|x)p(x). \end{aligned}$$

2.Expectations and covariance 期望和协方差

- 在知道一个函数自变量取值的概率分布情况下，可以求该函数的期望
  - $\mathbb{E}[f] = \sum_x p(x)f(x)$
  - $\mathbb{E}[f] = \int p(x)f(x)\,\mathrm{d}x.$
- 方差
  - $\mathrm{var}[f] = \mathbb{E}\left[(f(x) - \mathbb{E}[f(x)])^2\right]$
  - $E[(f(x)-E[f(x)])^2] = E\left[f(x)^2 - 2f(x)E\big(f(x)\big) + E[f(x)]^2\right]$
    $= E[f(x)^2] - 2E[f(x)]^2 + E[f(x)]^2 = E[f(x)^2] - E[f(x)]^2$