模式识别与机器学习

03 线性模型

大连理工大学 人工智能学院
School of Artificial Intelligence, Dalian University of Technology

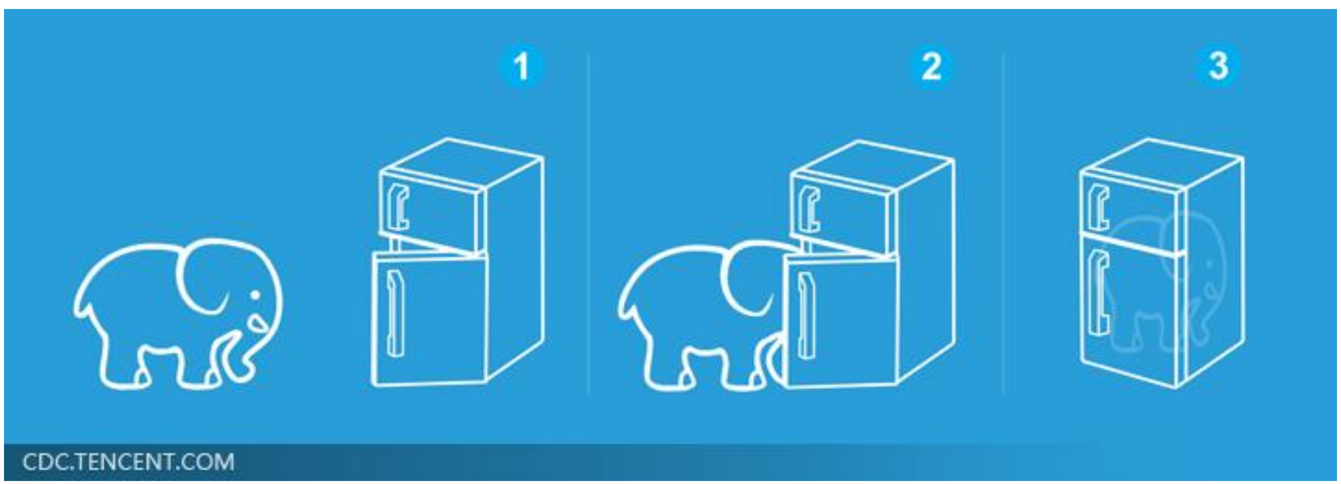
目录

- 线性回归
  - 最小二乘法
- 二分类任务
  - 对数几率回归
  - 线性判别分析
- 多分类学习
  - 一对一
  - 一对其余
  - 多对多
- 类别不平衡问题

机器学习

Step 1: define a set of function → Step 2: goodness of function → Step 3: pick the best function

就好像把大象放进冰箱……



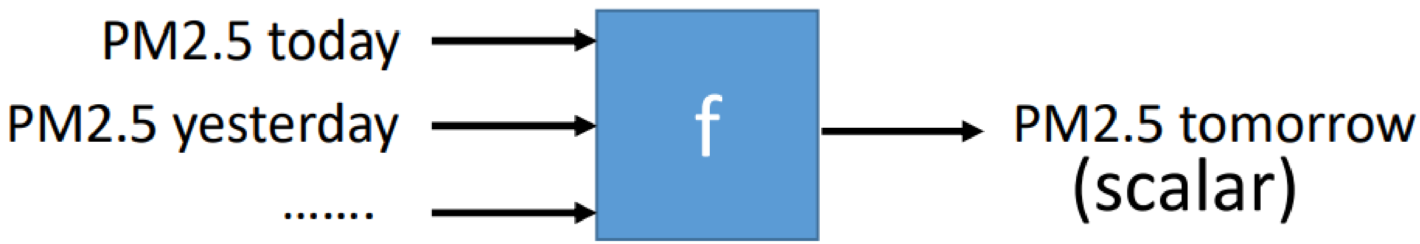
Credit： slide by Hung-yi Lee

机器学习

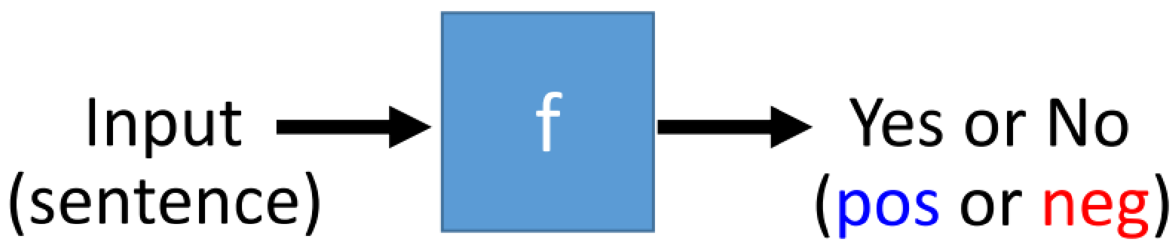
Step 1: define a set of function → Step 2: goodness of function → Step 3: pick the best function

Error

Regression



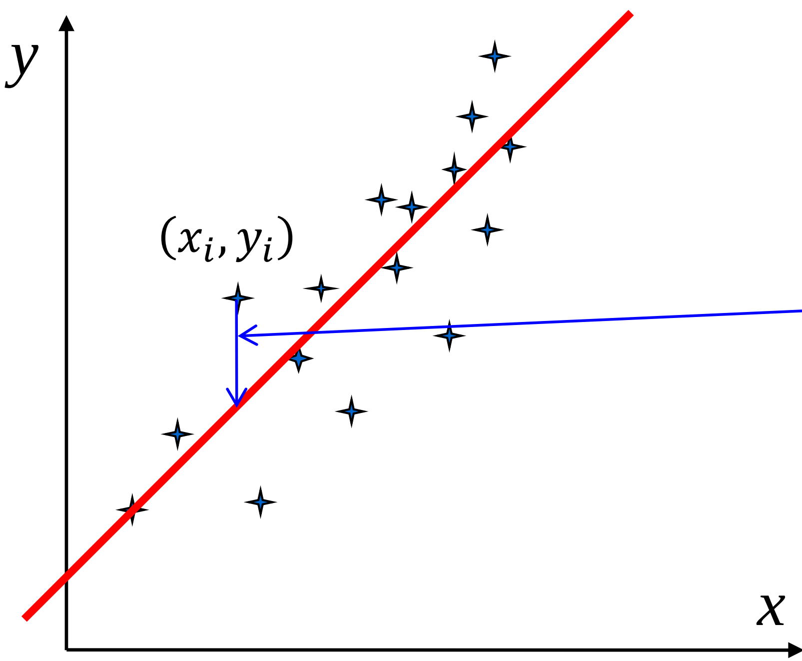
Classification



Different Tasks (任務)

Step 0: What kind of function do you want to find?

Step 1: define a set of function → Step 2: goodness of function → Step 3: pick the best function



Step 1:

$$f(x) = wx + b$$

Step 2: 回归误差

$$e_i = y_i - (wx_i + b)$$

Step 3:

$$\boxed{(w^*, b^*) = \arg\min_{w,b} \sum_{i=1}^{m} (y_i - wx_i - b)^2}$$

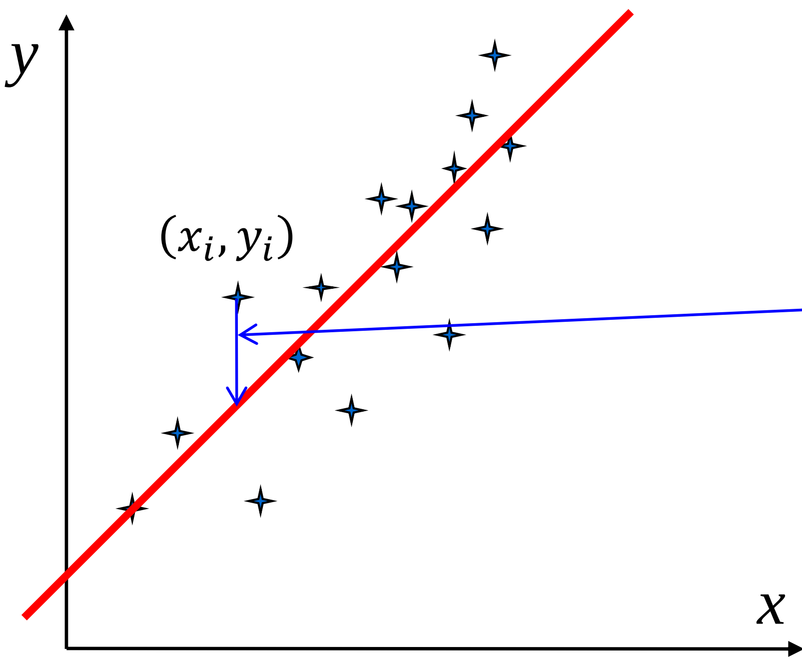
线性回归

线性回归

- 给定数据集 $D = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots, (\boldsymbol{x}_m, y_m)\}$

其中 $\boldsymbol{x}_i = (x_{i1}; x_{i2}; \ldots; x_{id})$, $y_i \in \mathbb{R}$

- 线性回归（linear regression）目的
  - 学得一个线性模型以尽可能准确地预测实值输出标记



Step 1:

$$f(x) = wx + b$$

Step 2: 回归误差

$$e_i = y_i - (wx_i + b)$$

Step 3:

$$\boxed{(w^*, b^*) = \arg\min_{w,b} \sum_{i=1}^{m} (y_i - wx_i - b)^2}$$

线性回归

- 单一属性的线性回归目标

$$f(x) = wx_i + b \text{ 使得 } f(x_i) \simeq y_i$$

- 参数/模型估计：最小二乘法（least square method）

$$\begin{aligned}(w^*, b^*) &= \underset{(w,b)}{\arg\min} \sum_{i=1}^{m} \left(f(x_i) - y_i\right)^2 \\ &= \underset{(w,b)}{\arg\min} \sum_{i=1}^{m} \left(y_i - wx_i - b\right)^2\end{aligned}$$

线性回归 - 最小二乘法

- 最小化均方误差

$$E_{(w,b)} = \sum_{i=1}^{m} (y_i - wx_i - b)^2$$

- 分别对 w 和 b 求导，可得

$$\frac{\partial E_{(w,b)}}{\partial w} = 2\left(w\sum_{i=1}^{m} x_i^2 - \sum_{i=1}^{m} (y_i - b)\, x_i\right)$$

$$\frac{\partial E_{(w,b)}}{\partial b} = 2\left(mb - \sum_{i=1}^{m} (y_i - wx_i)\right)$$

线性回归 – 最小二乘法

- 得到闭式（closed-form）解

$$w = \frac{\sum_{i=1}^{m} y_i (x_i - \bar{x})}{\sum_{i=1}^{m} x_i^2 - \frac{1}{m}\left(\sum_{i=1}^{m} x_i\right)^2}$$

$$b = \frac{1}{m}\sum_{i=1}^{m}(y_i - w x_i)$$

其中

$$\bar{x} = \frac{1}{m}\sum_{i=1}^{m} x_i$$

一般形式

- 线性模型一般形式

$$f(\boldsymbol{x}) = w_1 x_1 + w_2 x_2 + \ldots + w_d x_d + b$$

$\boldsymbol{x} = (x_1; x_2; \ldots; x_d)$是由属性描述的示例，其中$x_i$是$\boldsymbol{x}$在第$i$个属性上的取值

- 向量形式

$$f(\boldsymbol{x}) = \boldsymbol{w}^{\mathrm{T}} \boldsymbol{x} + b$$

其中 $\boldsymbol{w} = (w_1; w_2; \ldots; w_d)$

多元线性回归

- 给定数据集

$$D=\{(\boldsymbol{x}_1,y_1),(\boldsymbol{x}_2,y_2),\ldots,(\boldsymbol{x}_m,y_m)\}$$

$$\boldsymbol{x}_i=(x_{i1};x_{i2};\ldots;x_{id}) \quad y_i\in\mathbb{R}$$

- 多元线性回归目标

$$f(\boldsymbol{x}_i)=\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b \text{ 使得 } f(\boldsymbol{x}_i)\simeq y_i$$

多元线性回归

- 把$\boldsymbol{w}$和 b 吸收入向量形式 $\hat{\boldsymbol{w}} = (\boldsymbol{w}; b)$，数据集表示为

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1d} & 1 \\ x_{21} & x_{22} & \cdots & x_{2d} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{md} & 1 \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_1^{\mathrm{T}} & 1 \\ \boldsymbol{x}_2^{\mathrm{T}} & 1 \\ \vdots & \vdots \\ \boldsymbol{x}_m^{\mathrm{T}} & 1 \end{pmatrix}$$

$$\boldsymbol{y} = (y_1; y_2; \ldots; y_m)$$

多元线性回归 - 最小二乘法

- 最小二乘法（least square method）

$$\hat{\boldsymbol{w}}^* = \arg\min_{\hat{\boldsymbol{w}}} (\boldsymbol{y} - \mathbf{X}\hat{\boldsymbol{w}})^{\mathrm{T}} (\boldsymbol{y} - \mathbf{X}\hat{\boldsymbol{w}})$$

令 $E_{\hat{\boldsymbol{w}}} = (\boldsymbol{y} - \mathbf{X}\hat{\boldsymbol{w}})^{\mathrm{T}} (\boldsymbol{y} - \mathbf{X}\hat{\boldsymbol{w}})$ ，对$\hat{\boldsymbol{w}}$求导得到

$$\frac{\partial E_{\hat{\boldsymbol{w}}}}{\partial \hat{\boldsymbol{w}}} = 2\mathbf{X}^{\mathrm{T}} (\mathbf{X}\hat{\boldsymbol{w}} - \boldsymbol{y})$$

Matrix Cookbook
公式69和81

令上式为零可得$\hat{\boldsymbol{w}}$最优解的闭式解

多元线性回归 - 满秩讨论

- $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ 是满秩矩阵或正定矩阵，则

$$\hat{\boldsymbol{w}}^{*} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{y}$$

其中$\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}$是 $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ 的逆矩阵，线性回归模型为

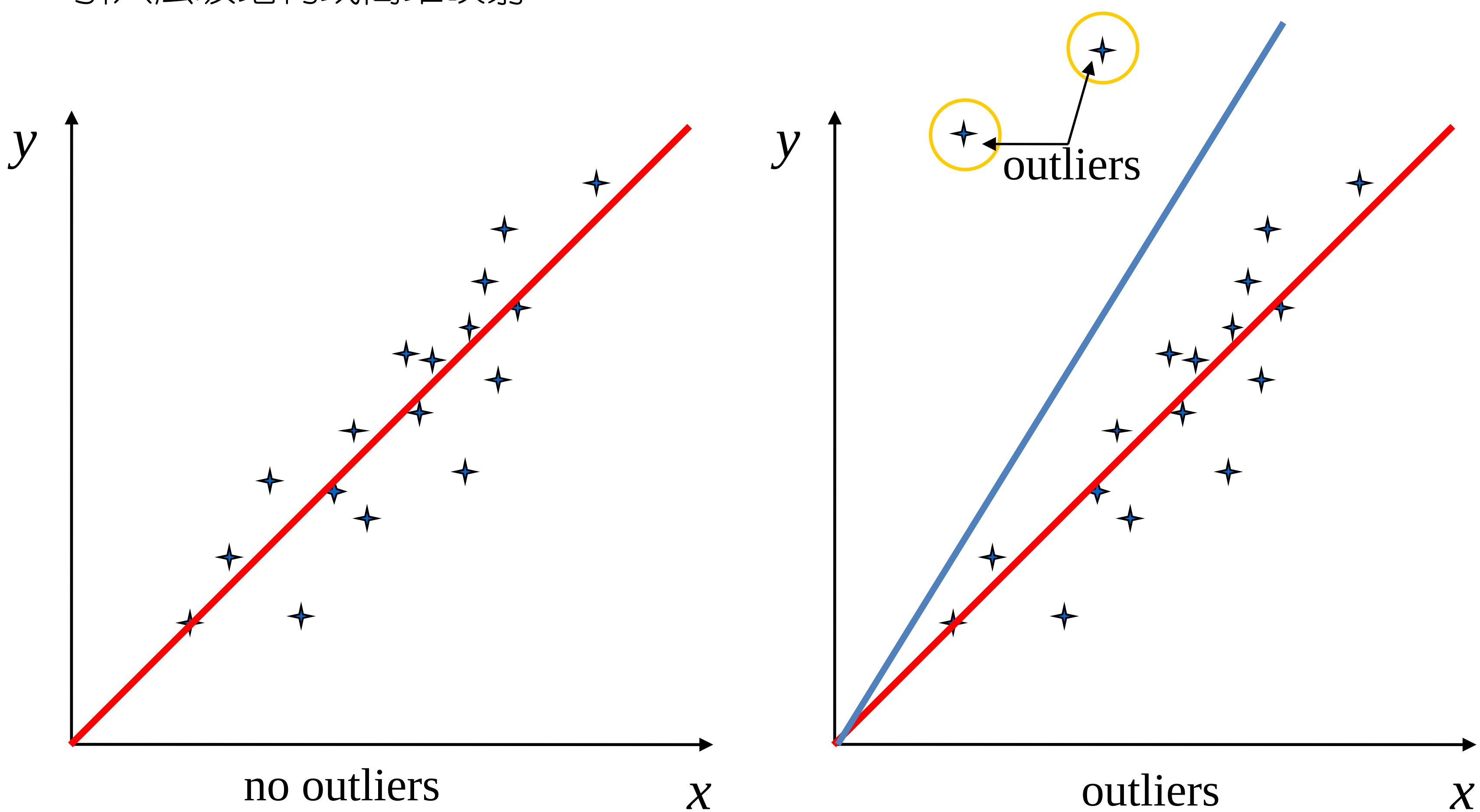
$$f\left(\hat{\boldsymbol{x}}_i\right) = \hat{\boldsymbol{x}}_i^{\mathrm{T}}\boxed{\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}}\mathbf{X}^{\mathrm{T}}\boldsymbol{y}$$

- $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ 不是满秩矩阵，多个解，怎么选？
  - 根据归纳偏好选择解（参见1.4节）
  - 引入正则化：对解空间的一种限制

- 正则化参考资料
- 吴恩达《机器学习》-正则化 https://www.bilibili.com/video/av55276229
- L1&L2正则化详解 https://www.bilibili.com/video/av77106463?from=search&seid=4369320229005019988
- 什么是L1 L2正则化? https://www.bilibili.com/video/av16009446?from=search&seid=4369320229005019988

线性模型特点

- 形式简单、易于建模
- 可解释性
- 非线性模型的基础
  - 引入层级结构或高维映射
- 对异常点鲁棒性差
  - 随机取样一致 (Random Sample Consensus, RANSAC)
  - 鲁棒回归 (Robust Regression)



线性的含义

$$y = wx + b$$ 线性？

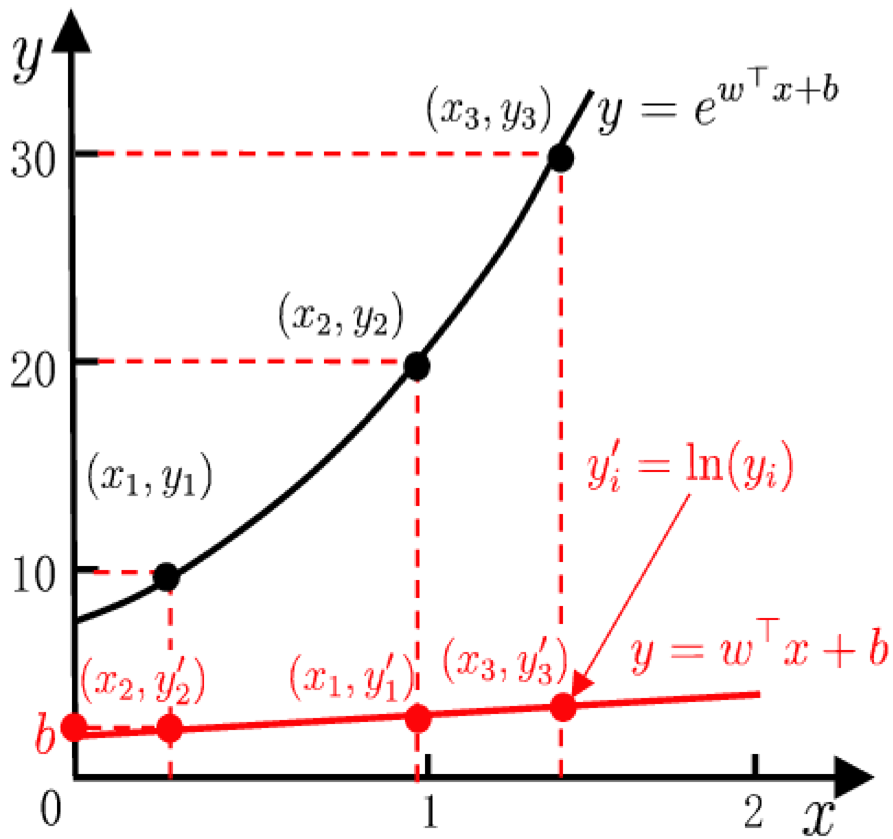
$$y = w_1 x + w_2 x^2 + w_3 x^3 + b$$ 线性？

$$x_1 = x,\ x_2 = x^2,\ x_3 = x^3$$
$$y = w_1 x_1 + w_2 x_2 + w_3 x_3 + b$$ 线性？

- 线性并不指对输入变量的线性，而是指对参数空间的线性。也就说对于输入来说，完全可以对先对其进行非线性变换，再进行线性组合。从这个角度来说，线性模型完全具有描述非线性的能力。
- 通用非线性化方法：核学习方法 (Kernel-based Learning Algorithms)

对数线性回归

- 输出标记的对数为线性模型逼近的目标



$$\ln y = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x} + b$$

$$y = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x} + b$$

广义线性模型

回归问题的最简单模型

$$y(\boldsymbol{x}, \boldsymbol{w}) = w_0 + w_1 x_1 + \ldots + w_D x_D$$ 偏置参数

扩展模型类别

$$y(\boldsymbol{x}, \boldsymbol{w}) = w_0 + \sum_{j=1}^{M-1} w_j \phi_j(\boldsymbol{x})$$ 基函数

令 $\phi_0(\boldsymbol{x}) = 1$

$$y(\boldsymbol{x}, \boldsymbol{w}) = \sum_{j=0}^{M-1} w_j \phi_j(\boldsymbol{x}) = \boldsymbol{w}^T \boldsymbol{\phi}(\boldsymbol{x})$$

线性模型

$$\boldsymbol{w} = (w_0, \ldots, w_{M-1})^T \quad \boldsymbol{\phi} = (\phi_0, \ldots, \phi_{M-1})^T$$

表示特征

使用非线性基函数，能够使函数成为输入向量 x 的一个非线性函数