NN_NLP CHAPTER 3

From Linear Models to Multi-layer Perceptions

马新宇、刘嘉铭

2018/9/17

Outline

- 线性模型的问题 （chapter3.1）
  - 无法解决异或问题
- 如何解决？（chapter3.2-4）
  - 核技巧
  - 映射函数

Limitations of linear models:XOR

$$\mathrm{xor}(0, 0) = 0$$
$$\mathrm{xor}(1, 0) = 1$$
$$\mathrm{xor}(0, 1) = 1$$
$$\mathrm{xor}(1, 1) = 0$$

线性分类器: $f(x) = xw + b$

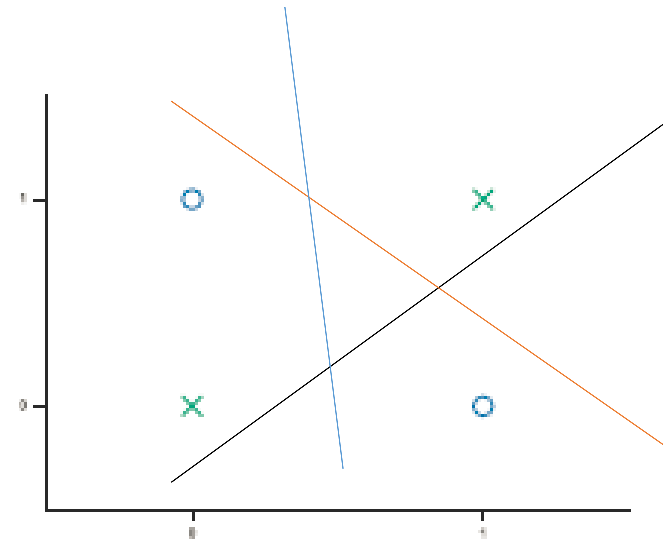
$(0,0)w + b < 0$

$(0,1)w + b \geq 0$

$(1,0)w + b \geq 0$

$(1,1)w + b < 0$

无解！

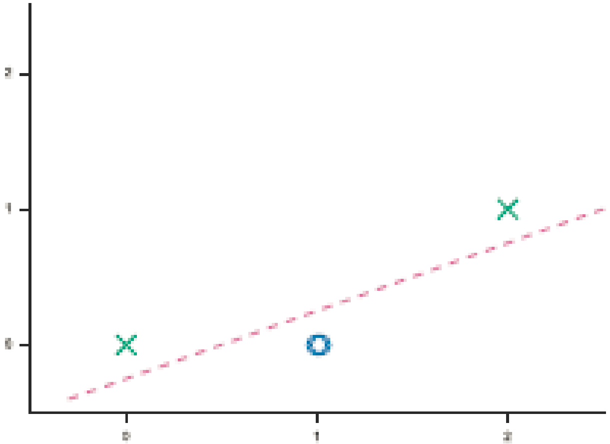


How to solve non-linearly separable

将数据映射为适合线性分类的表示: $x \rightarrow \varphi(x)$

$$\hat{y} = f(x) = \varphi(x)w + b$$

通常需要将数据映射到更高的维度



问题:

1>映射函数人工定义

2>且高度依赖数据集

两种方法:

1>核方法(手动定义)

2>可训练的映射函数

1.Kernel method

- Kernel定义：给定输入空间X和特征空间H (Hilbert Space)，如果存在一个映射从X到H，对于任何$x, z \in X$，函数$K(x, z)$满足$K(x, z) = <\varphi(x), \varphi(z)>$，那么$K(x, z)$称为核，$\varphi$是映射函数。
- 希尔伯特空间：完备的内积空间。
- 线性空间引入内积来衡量向量角度的空间叫内积空间。
- 完备性：在该空间内所有的运算结果也在该空间内，即极限不会超出这个空间。

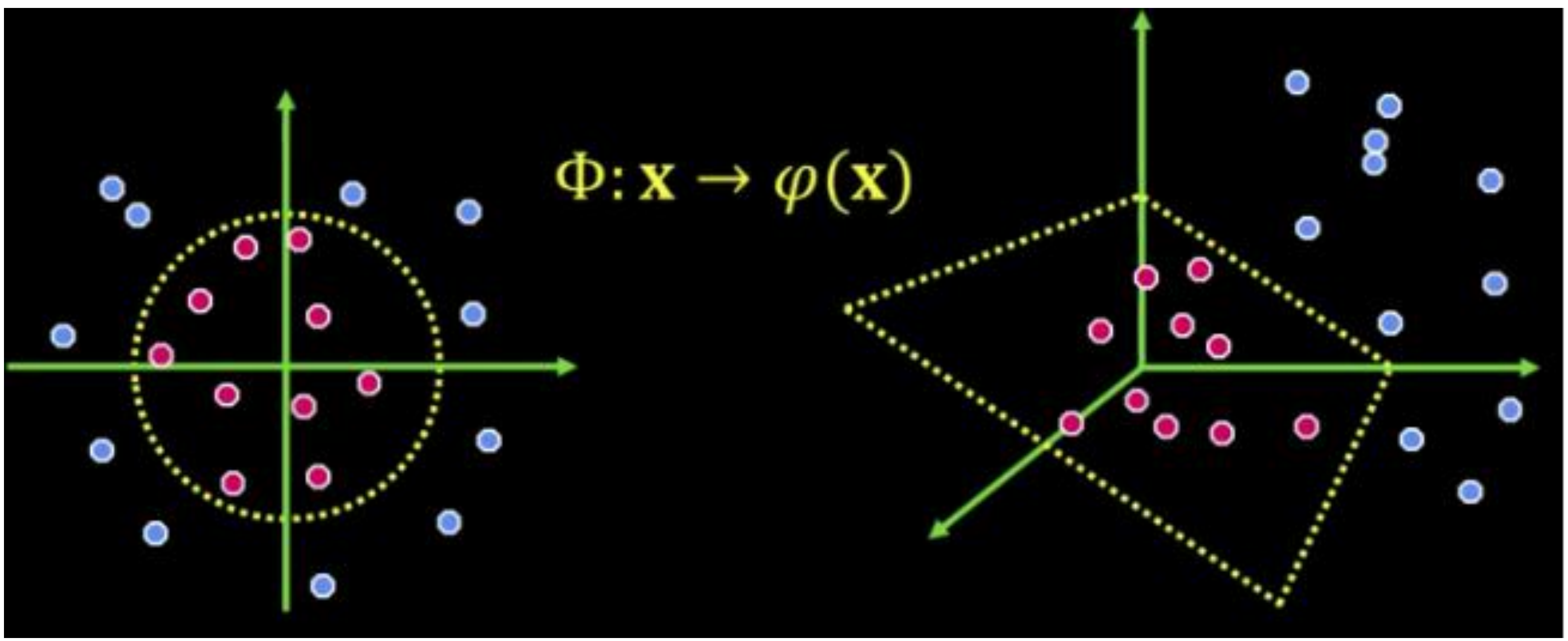
1.Kernel method

- 对于一个给定的核$K(x,z)$，特征空间H 和映射函数通常是不唯一的。
- 例如：$X=R^2, K(x,z) = <x,z>^2$, $\mathrm{x}=(x_1,x_2)$, $\mathrm{z}=(z_1,z_2)$
- $<x,z>^2 = (x_1\mathrm{z}_1 + x_2 z_2)^2$
  $= x_1{\mathrm{z}_1}^2 + 2x_1\mathrm{z}_1 x_2 z_2 + x_2{\mathrm{z}_2}^2$

$$H=R^3,\ \phi(x)=(x_1^2,\sqrt{2}x_1x_2,x_2^2)^T.$$
$$H=R^3,\ \phi(x)=\frac{1}{\sqrt{2}}((x_1-x_2)^2,2x_1x_2,(x_1+x_2)^2)^T.$$
$$H=R^4,\ \phi(x)=(x_1^2,x_1x_2,x_1x_2,x_2^2)^T.$$

1.Kernel method

- 将线性不可分数据映射到更高维空间寻找可分
- 但在高维空间中计算过于复杂，φ难定义



1.Kernel method

- TRICK：我们不需要显式的去定义映射函数φ，也不需要计算经过转换后的高维数据，只需计算内积也即核函数$K(x,z)$就可以。
- SVM对偶形式的目标函数：

$$\max_{\alpha} \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{n} y^{(i)}y^{(j)}\alpha_i\alpha_j (x^{(i)})^T x^{(j)}$$

$$s.t. \quad \alpha_i \geq 0, i = 1, \cdots, n,$$

$$\sum_{i=1}^{n}\alpha_i y^{(i)} = 0.$$

1.Kernel method

- 常用的核函数：多项式核、高斯核
- 多项式核：$K(x,y)=(x^T y)^2$
- 高斯核：$K(x,y)=\exp(-\|x-y\|)^2$
- 多项式核函数将低维数据映射到高维(维度是有限的)，那么对于(无限个不同维的多项式核函数之和)高斯核，其维度是无限的。
- 缺点：
  - SVM分类过程线性依赖于训练集大小
  - 增加了过拟合风险

2.Trainble mapping function

- 定义一个可训练的映射函数和分类器同时训练
- $\hat{y} = f(x) = \ \varphi(x)w + b$
- $\varphi(x) = g(xW' + b')$，g函数是非线性的

- 同时学习“表示函数”和在其之上的分类器也是神经网络背后的主要想法
- 上式也描述了最基本的神经网络——感知机