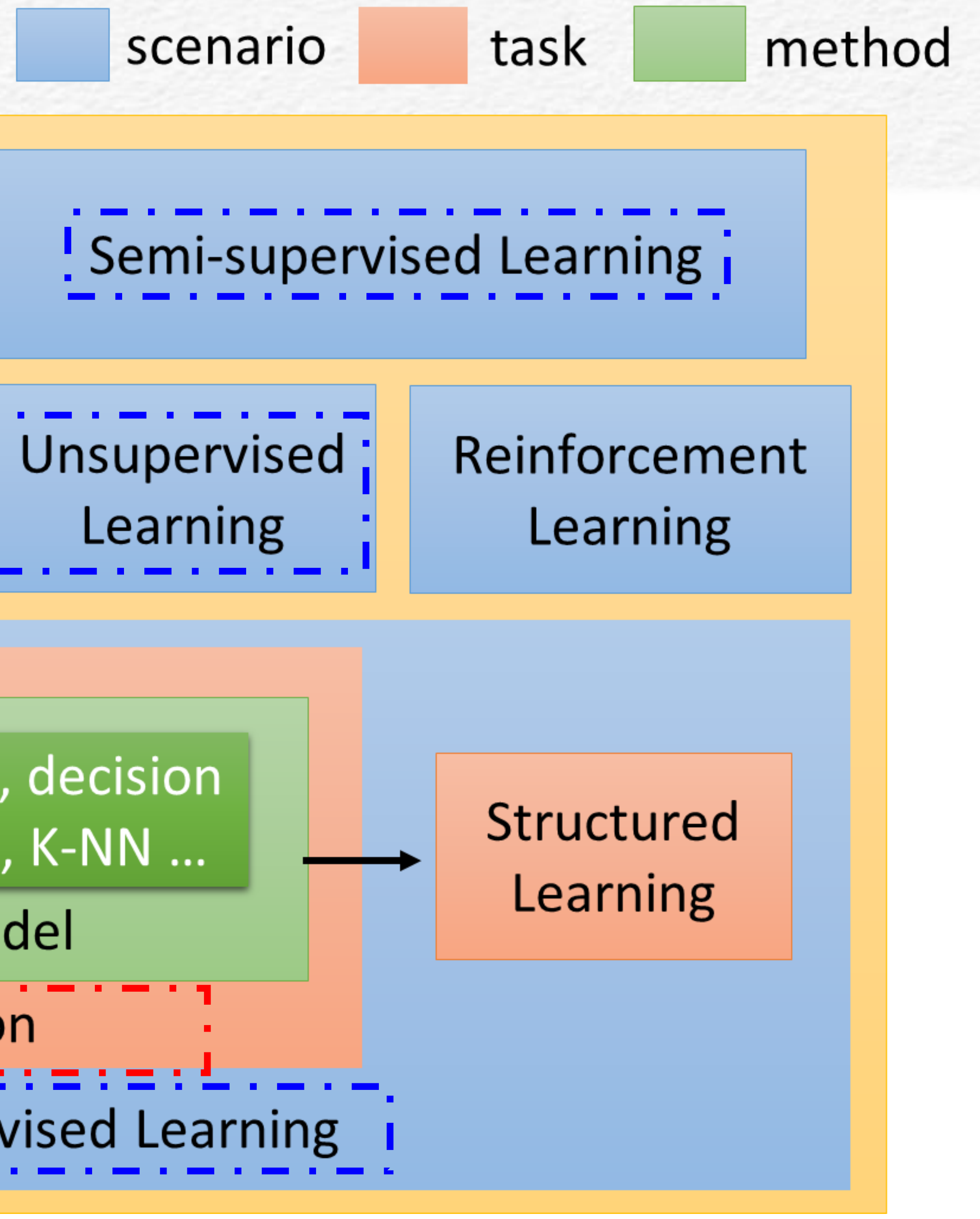
期中复习

2025年4月11日

期中考试

- 时间：4月18日 9:55-11:30
- 题型：
  - 填空（30分）
  - 简单（10分）
  - 综合题（60分）
- 闭卷，允许带计算器
- 考试范围：
  - 以下ppt提到的相关内容

机器学习


scenario
task
method
Learning Theory
Regression
Linear Model
Neural network
SVM, decision tree, K-NN …
Non-linear Model
Classification
Supervised Learning
Semi-supervised Learning
Unsupervised Learning
Reinforcement Learning
Structured Learning

应用

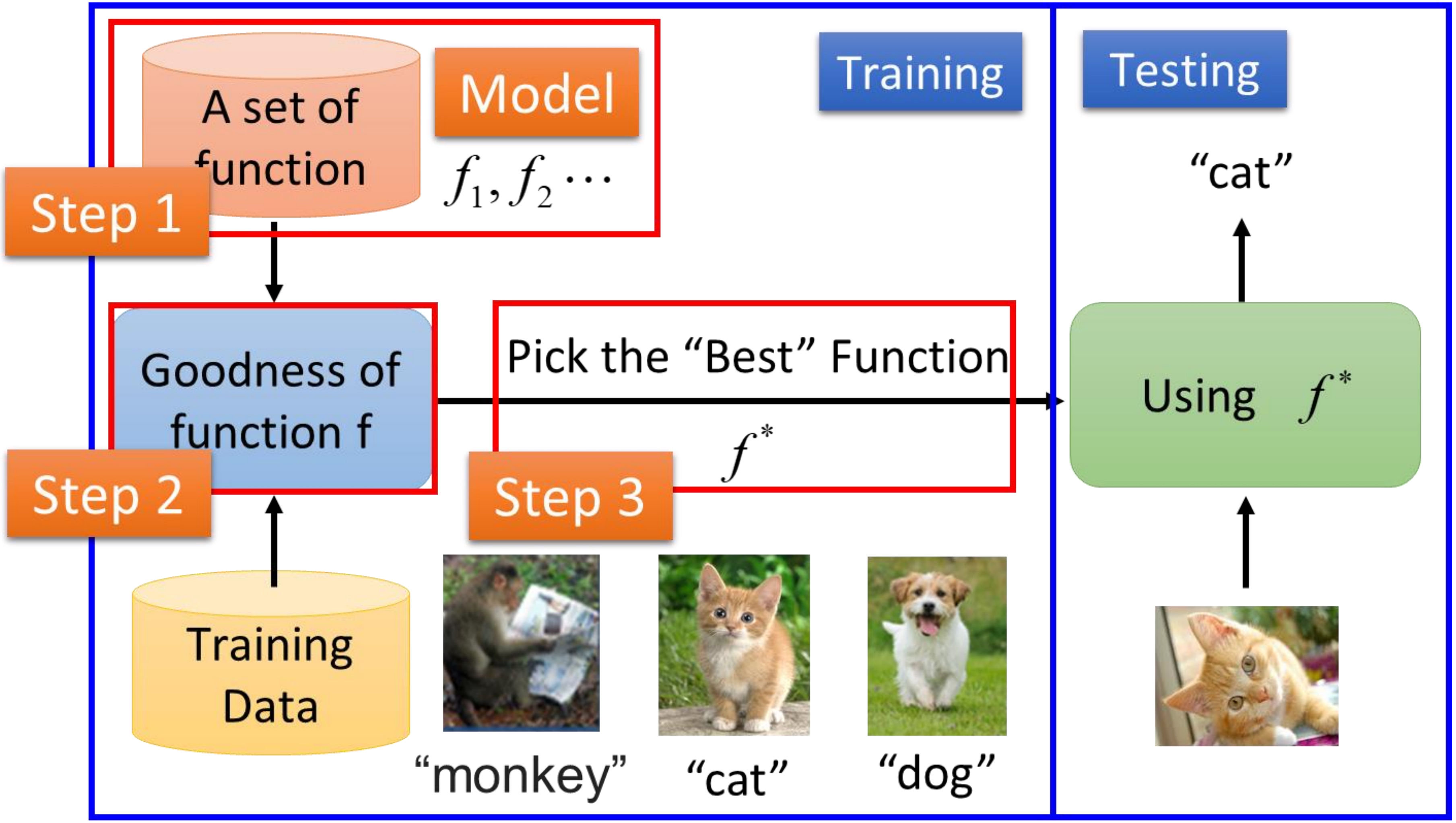
回归/预测 →

路段流量预测
空气质量预测
生成产量预测
股票、房价预测
……

分类 →

文本分类
情感分类
图像分类
序列标注
……

机器学习



Training
Testing
A set of function
Model
$f_1, f_2 \cdots$
Step 1
"cat"
Goodness of function f
Pick the "Best" Function
f^*
Using f^*
Step 2
Step 3
Training Data
"monkey"
"cat"
"dog"

基本要素

- 方法 = 模型 + 策略 + 算法
- 模型：学习什么样的模型？
- 策略：按照什么准则学习或选择最优的模型？
  - 定义损失函数：
    - $L = \frac{1}{N}\sum_{i=1}^{N}(f(x_i) - y_i)^2$
    - $L = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{C} y_{i,k}\log f_k(x_i)$
- 算法：学习模型的具体计算方法

基本要素

- 算法：学习模型的具体计算方法
  - 梯度下降 $w^*, b^* = arg\min_{w,b} L(w,b)$
    - ➢ (Randomly) Pick an initial value w^0, b^0
    - ➢ Compute $\frac{\partial L}{\partial w}|_{w=w^0,b=b^0}$, $\frac{\partial L}{\partial b}|_{w=w^0,b=b^0}$

$$w^1 \leftarrow w^0 - \eta \frac{\partial L}{\partial w}|_{w=w^0,b=b^0} \qquad b^1 \leftarrow b^0 - \eta \frac{\partial L}{\partial b}|_{w=w^0,b=b^0}$$

    - ➢ Compute $\frac{\partial L}{\partial w}|_{w=w^1,b=b^1}$, $\frac{\partial L}{\partial b}|_{w=w^1,b=b^1}$

$$w^2 \leftarrow w^1 - \eta \frac{\partial L}{\partial w}|_{w=w^1,b=b^1} \qquad b^2 \leftarrow b^1 - \eta \frac{\partial L}{\partial b}|_{w=w^1,b=b^1}$$

过拟合

- 模型对已知数据预测很好，对未知数据预测很差
  - 增加训练集数量
  - 正则化

l2范式：$\|\boldsymbol{w}\|_2^2 = \boldsymbol{w}_1^2 + \boldsymbol{w}_2^2 + \cdots + \boldsymbol{w}_d^2$

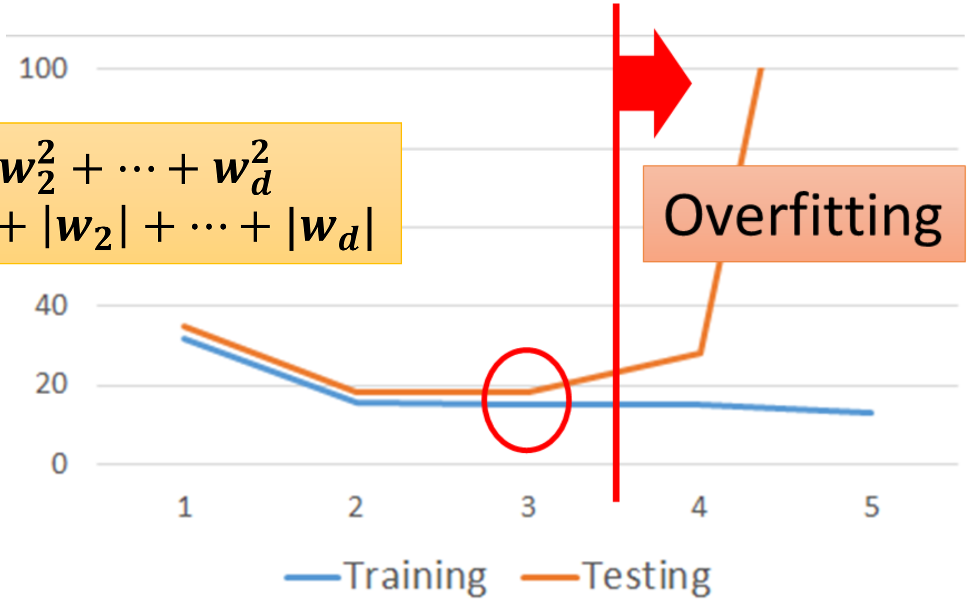
l1范式：$\|\boldsymbol{w}\|_1 = |\boldsymbol{w}_1| + |\boldsymbol{w}_2| + \cdots + |\boldsymbol{w}_d|$

$$L(w) = \frac{1}{2}\sum_{i=1}^{N}(f(x_i, w) - y_i)^2 + \frac{\lambda}{2}||w||_2^2$$

$$L(w) = \frac{1}{2}\sum_{i=1}^{N}(f(x_i, w) - y_i)^2 + \lambda||w||_1$$

控制模型预测准确性

控制模型复杂度



Model

线性回归

感知机

逻辑回归

支持向量机

K近邻&聚类

PCA

贝叶斯方法

Model

线性回归

$$Y = w^T X + b$$

$Y \in R^{1\times N}$ $w \in R^{d\times 1}$

$X \in R^{d\times N}$ $b \in R^{1\times N}$

$$L(w,b) = \sum_{i=1}^{N}\left(\left(w^T x^i + b\right) - y^i\right)^2$$

梯度下降

Model

感知机

决策函数

$$f(x) = sign(w^T x + b)$$

线性可分数据集

输入：特征向量

输出：实例的类别，取+1和−1二值

误分类点驱动

误分类点到超平面的距离

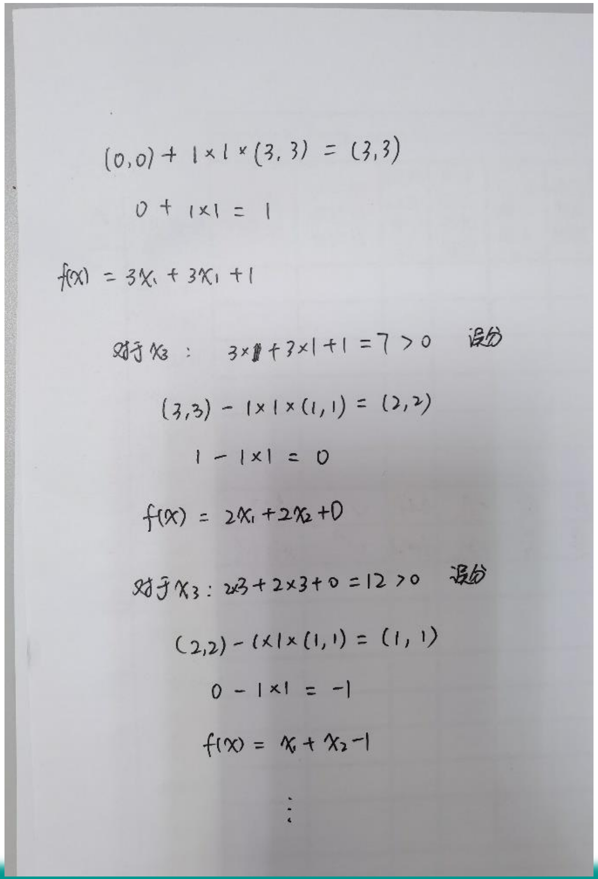
$$L(w,b) = -\sum_{x_i \in M} y_i(w^T x_i + b)$$

随机梯度下降

$$\boldsymbol{y_i(w^T x_i + b) \le 0}$$

感知机学习算法

- 已知训练数据集，正实例点是$x_1=(3,3)^T$, $x_2=(4,2)^T$, 负实例点是$x_3=(1,1)^T$, $x_4=(2,0)^T$，用原始形式求出感知机模型决策函数f=sign(wx+b)，按照例题形式给出过程



$(0,0)+1\times1\times(3,3)=(3,3)$

$0+1\times1=1$

$f(x)=3x_1+3x_1+1$

对于x_3： $3\times1+3\times1+1=7>0$ 误分

$(3,3)-1\times1\times(1,1)=(2,2)$

$1-1\times1=0$

$f(x)=2x_1+2x_2+0$

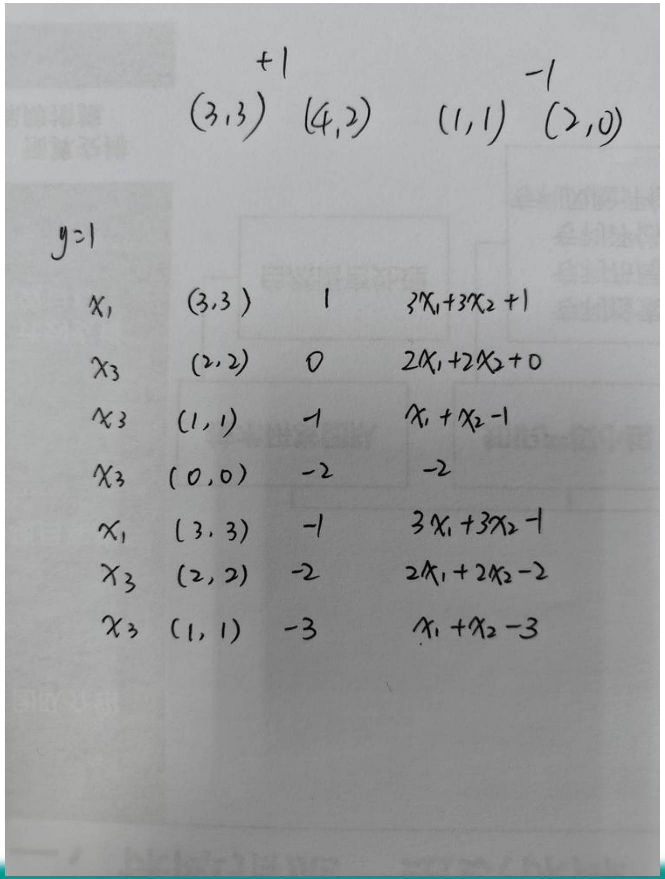
对于x_3：$2\times3+2\times3+0=12>0$ 误分

$(2,2)-1\times1\times(1,1)=(1,1)$

$0-1\times1=-1$

$f(x)=x_1+x_2-1$

⋮



+1 $(3,3)$ $(4,2)$ -1 $(1,1)$ $(2,0)$

$y=1$

x_1	(3,3)	1	$3x_1+3x_2+1$
x_3	(2,2)	0	$2x_1+2x_2+0$
x_3	(1,1)	-1	x_1+x_2-1
x_3	(0,0)	-2	-2
x_1	(3,3)	-1	$3x_1+3x_2-1$
x_3	(2,2)	-2	$2x_1+2x_2-2$
x_3	(1,1)	-3	x_1+x_2-3

Model

~~线性~~回归

逻辑

$$Y = w^T X + b$$

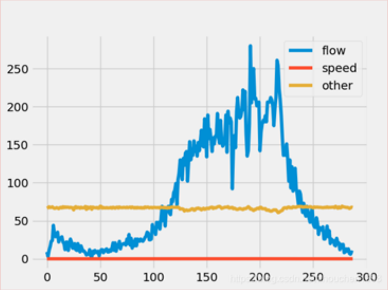
$$f_{w,b}(X) = \frac{1}{1 + exp(-(w^T X + b))}$$

$$L(w,b) = \sum_{i=1}^{N} \left((w^T x^i + b) - y^i \right)^2$$

$$L(w,b) = \sum_{n} -\left[y^i \ln f_{w,b}(x^i) + (1 - y^i) ln \left(1 - f_{w,b}(x^i) \right) \right]$$

梯度下降

Model

	感知机	逻辑回归	线性回归
训练数据格式	Dog recognition  1  -1	Dog recognition  1  0	
决策函数	$f = sign(wx + b)$	$f = \frac{1}{1 + e^{-(wx+b)}}$	$y = wx + b$
损失函数	$l = -\sum_{x_i \in M} y_i(wx_i + b)$	$l = -\sum_{i=1}^{N}[y_i \ln f(x_i) + (1 - y_i)\ln(1 - f(x_i))]$	$l = \sum_{i=1}^{N}[f(x_i) - y_i]^2$
训练结果	模型输出为+1/-1 直接给出类别标签	模型输出为(0,1) 根据阈值0.5进行判断，大于0.5属于类别1，小于0.5属于类别0	模型输出预测值

Model

支持向量机

决策函数

$f(x) = sign(w^T x + b)$

通过寻求**结构化风险最小**来提高学习机的**泛化能力**

在特征空间上的**间隔最大**的线性分类器

$$\min_{w,b} \frac{1}{2}||w||^2$$

$$s.t.\ \ y_i(wx_i + b) - 1 \geq 0, \qquad i = 1,2,\cdots,N$$

$$L = \frac{1}{2}||w||^2 + \sum_{i=1}^{N} \alpha_i (1 - y^i (w \cdot x^i + b)),\ \ \alpha_i \geq 0$$

原问题 $\min_{x} \max_{\alpha,\beta} L(x,\alpha,\beta)$ $\leftrightarrow$ 对偶问题 $\max_{\alpha_i \geq 0} (\min_{b,w} L(w,b,\alpha))$

Model

支持向量机

$$\boldsymbol{L} = \frac{1}{2}||\boldsymbol{w}||^2 + \sum_{i=1}^{N} \boldsymbol{\alpha}_i \left(\mathbf{1} - \boldsymbol{y}^i \left(\boldsymbol{w} \cdot \boldsymbol{x}^i + \boldsymbol{b}\right)\right), \quad \boldsymbol{\alpha}_i \geq \mathbf{0}$$

对偶问题 $\max\limits_{\boldsymbol{\alpha}_i \geq 0}(\min\limits_{\boldsymbol{b},\boldsymbol{w}} \boldsymbol{L}(\boldsymbol{w}, \boldsymbol{b}, \boldsymbol{\alpha}))$

$$\frac{\partial L}{\partial w_j} = w_j - \sum_{i=1}^{N} \alpha_i y^i x_j^i = 0 \quad \boldsymbol{w}_j = \sum_{i=1}^{N} \boldsymbol{\alpha}_i \boldsymbol{y}^i \boldsymbol{x}_j^i$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{N} \boldsymbol{\alpha}_i \boldsymbol{y}^i = \mathbf{0}$$

$$\min_{\alpha} L = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j \, y^i y^j \left(\boldsymbol{x}^i \cdot \boldsymbol{x}^j\right) - \sum_{i=1}^{N} \alpha_i$$

$$\sum_{i=1}^{N} \alpha_i y^i = 0, \quad \alpha_i \geq 0$$

Model

支持向量机

线性可分支持向量量

线性支持向量量

$$\min_{w,b} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i$$

$$s.t.\ \ y_i(wx_i + b) \geq 1 - \xi_i\,,\quad i = 1,2,\cdots,N$$

非线性支持向量量

$$\min_{\alpha} \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^{N}\alpha_i$$

$$\text{s.t.}\ \ \sum_{i=1}^{N}\alpha_i y_i = 0$$

$$0 \leq \alpha_i \leq C$$

决策函数

$$f(x) = sign(w^T\Phi(x) + b)$$

$$f(x) = sign\left(\sum_{i=1}^{N}\alpha^* y_i K(x, x_i) + b^*\right)$$

支持向量机

- 求解

- 参见例题7.2
  - 正实例$\boldsymbol{x_1} = (\mathbf{3}，\mathbf{3})^{\boldsymbol{T}}, \boldsymbol{x_2} = (\mathbf{4}，\mathbf{3})^{\boldsymbol{T}}$, 负实例$\boldsymbol{x_3} = (\mathbf{1}，\mathbf{1})^{\boldsymbol{T}}$，求线性支持向量机

Model

K近邻

➢ KNN的学习过程

- Lazing learning
- 投票原则
- 距离度量

欧氏距离:

$$L_2(x_i, x_j) = \left(\sum_{l=1}^{d} \left| x_i^{(l)} - x_j^{(l)} \right|^2 \right)^{\frac{1}{2}}$$

Model

聚类

对大量未标注的数据集，按数据的内在相似性将数据集进行划分，形成多个簇，使得内部数据相似度尽可能大而类别间的数据相似度尽可能小

K-means

通过迭代把数据集划分为不同的类别，

使得评价聚类性能的准则函数达到最优，

使得每个聚类类内紧凑，类间独立

层次聚类

层次对数据集进行划分，从而形成树形的聚类结构

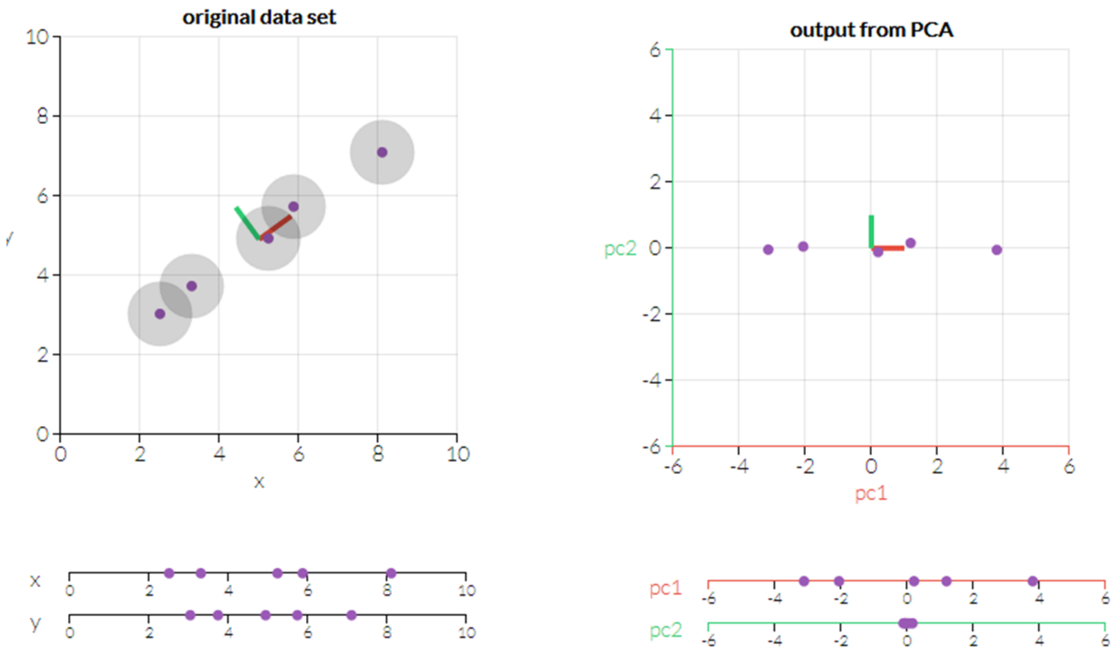
参见例题14.1

Model

PCA

用于高维数据的降维。提取数据的主要特征分量

通过正交变换将原始数据转换到新的坐标系中



最大可分性——方差最大

目标：方差最大化

通过选择方差最大的方向作为第一主成分

— 通过特征分解，找到特征向量

— 每一个特征向量均为一个一维基

Model

后验概率计算： $P(B_i|A)=\frac{P(A|B_i)|P(B_i)}{P(A)}=\frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{N}P(A|B_j)P(B_j)}$

贝叶斯方法

朴素贝叶斯

前提：贝叶斯定理和特征条件独立

$f(x)=argmaxP(c)\prod_{i=1}^{d}P(x_i|c)$

参见课堂例题，例4.1

贝叶斯网络

有向无环图+条件概率表

预测

推理

参见课堂例题，课堂作业（提交有答案）