模式识别引论

An Introduction to Pattern Recognition

主讲: 李春光

www.pris.net.cn/teacher/lichunguang

模式识别与智能系统实验室

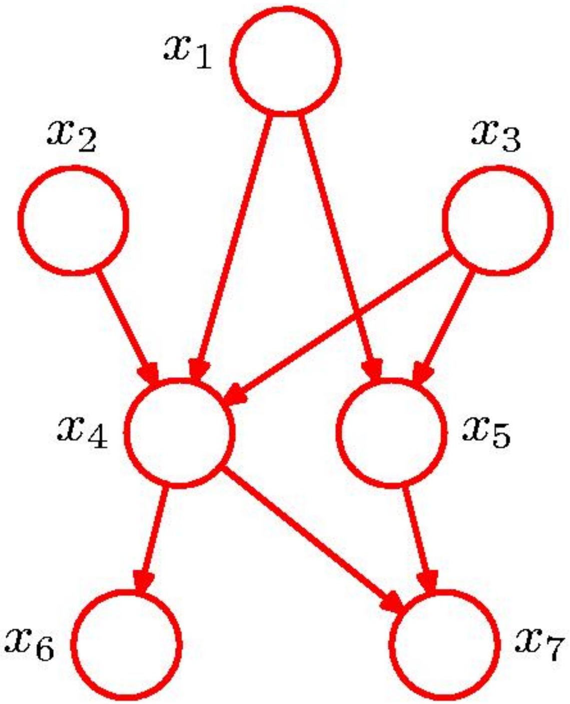
网络搜索教研中心 信息与通信工程学院 北京邮电大学

本学期内容总结

- 引言
  - 模式识别中的基本问题
- 部分数学基础
  - 概率论\决策论\信息论
  - 贝叶斯公式\最大似然估计\最大后验概率估计\贝叶斯估计\最大熵原理
- 线性模型for回归
  - 最小二乘法 \ 正则化的最小二乘法 \ 贝叶斯线性回归
  - 几种常用分布 \ 共轭先验
- 线性模型 for 分类
  - 感知器 \ **Logistic**回归 \ 迭代再加权最小二乘法 \ 贝叶斯**Logistic**回归
- 神经网络
  - 多层感知器 \ 反向传播(**BP**)算法
- 核方法
  - 对偶表示 \ 核的构造方法 \ 高斯过程
- 支持向量机(SVM)
  - 最大间隔分类 \ **SVM** \ 二次规划

- 概率图模型 \ 降维技术 \ 聚类算法 等

Bayesian Networks

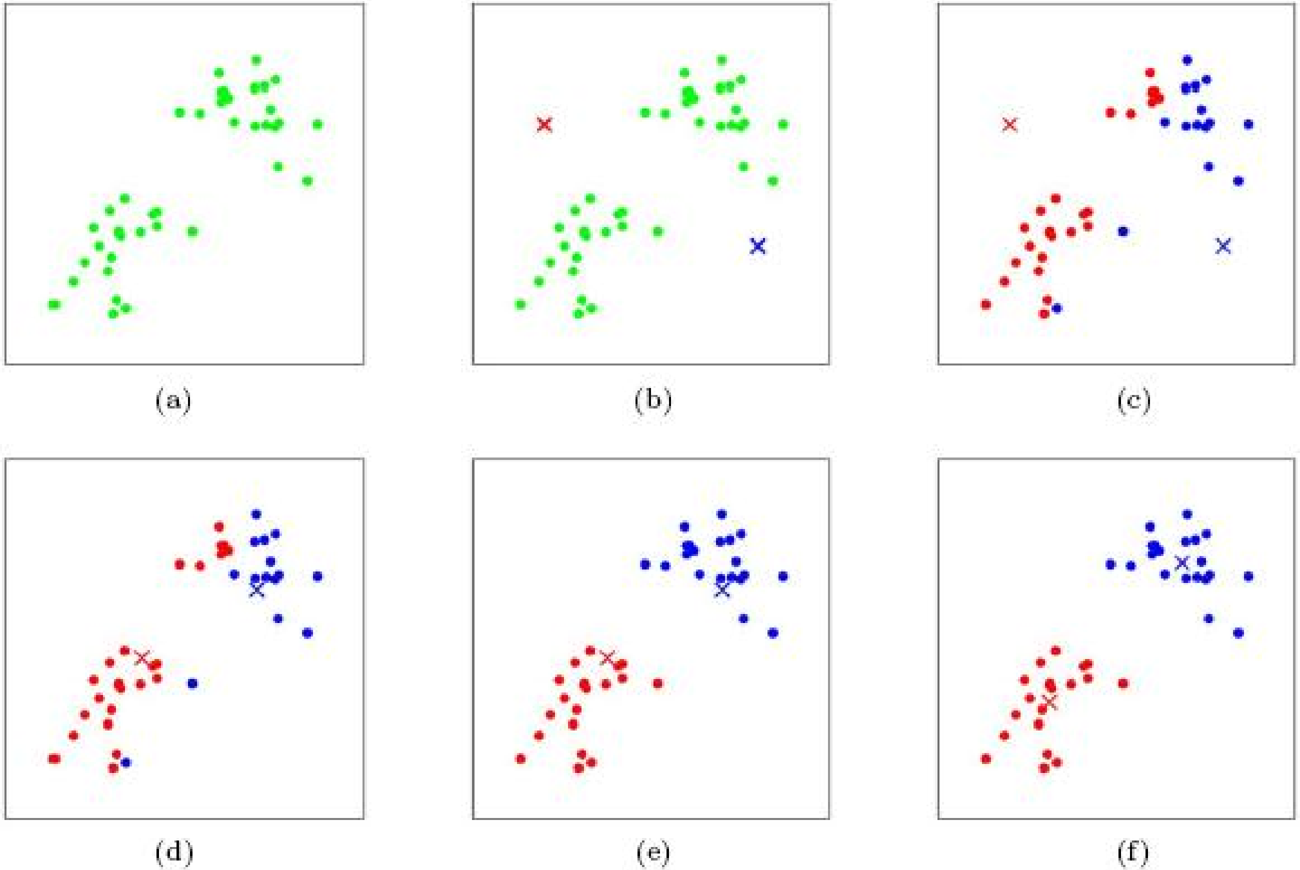


$$
\begin{aligned}
p(x_1, \ldots, x_7) \quad = \quad & p(x_1)p(x_2)p(x_3)p(x_4|x_1, x_2, x_3) \\
& p(x_5|x_1, x_3)p(x_6|x_4)p(x_7|x_4, x_5)
\end{aligned}
$$

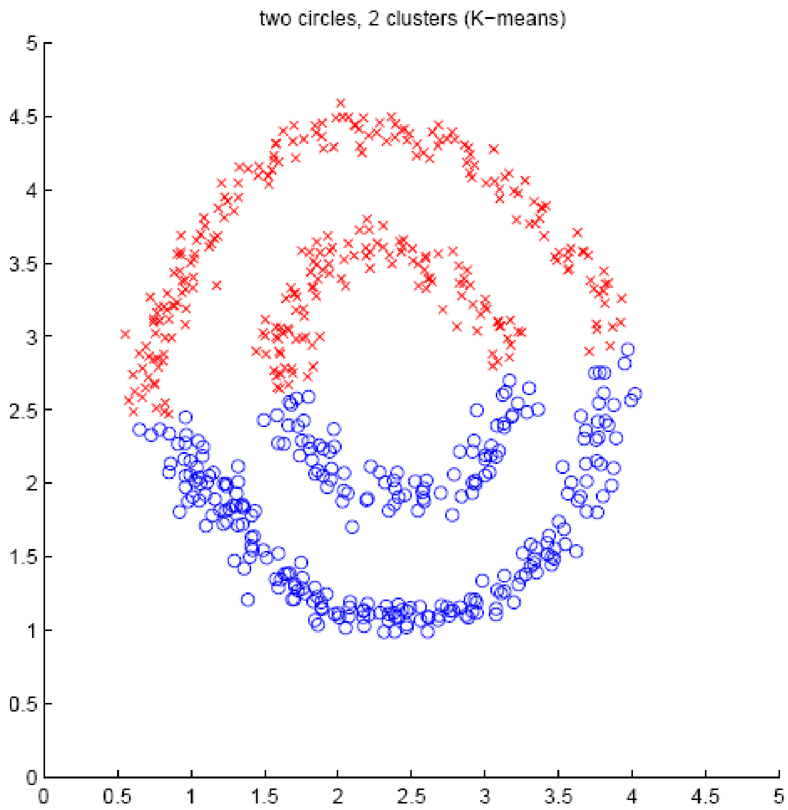
The absence of links conveys important information.

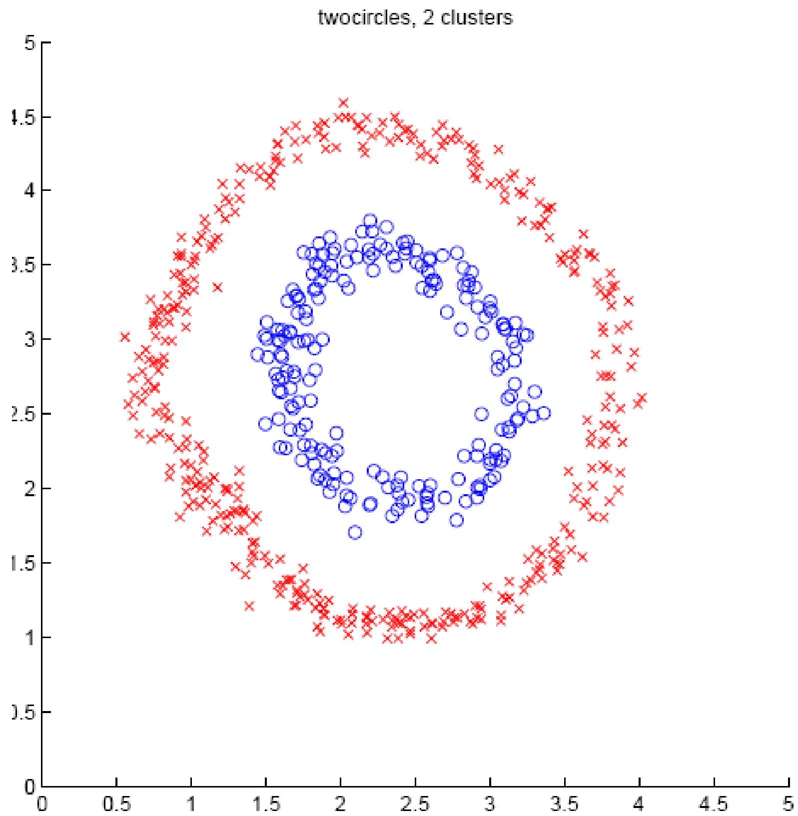
k-均值聚类(k-means clustering)

- k>1是预先指定的聚类数目
  - 比如 **k=2**



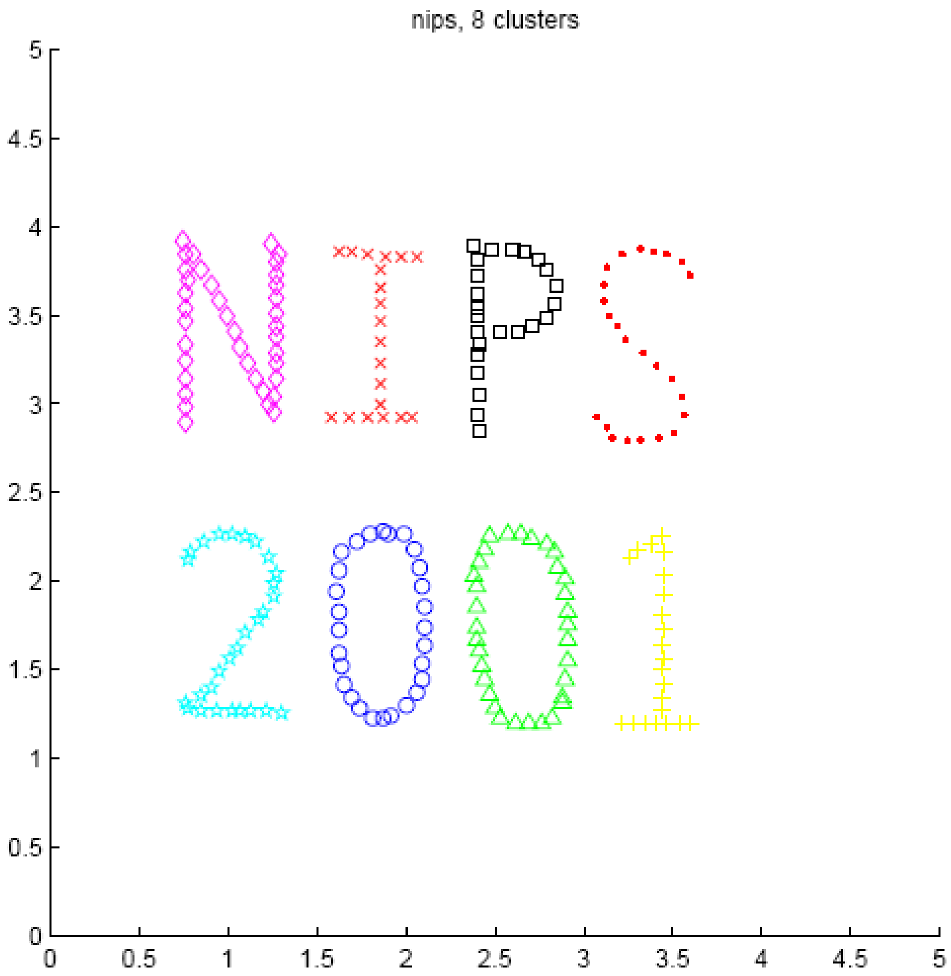
如果数据中的聚类结构是non-convex...





如果数据中的聚类结构是non-convex…

– **k-均值聚类 → 谱聚类(Spectral Clustering)**



拉普拉斯特征映射

- 理论来源
  - 特征图理论(Spectral Graph Theory)
- 定义目标函数
  - 保证相似的点在投影之后保持相似，不相似的点在投影之后相互远离

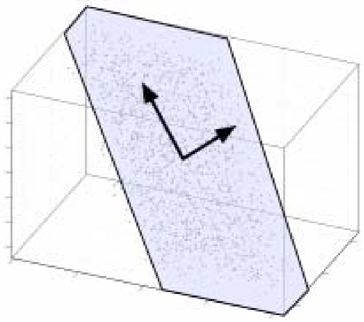
$$\arg\min_{Y}\sum_{i=1}^{N}\sum_{j=1}^{N}w_{ij}\left\|z_i-z_j\right\|^2=\arg\min_{Y} tr\left(Y^T LY\right)\quad s.t.\quad Y^T DY=I$$

- L = D － W： Laplacian of G(V,E,W)， $Y_{N\times d}=\left(\mathbf{y}_1\mathbf{y}_1\ldots\mathbf{y}_d\right)$,

$$w_{ij}=\exp\left(-\left\|\mathbf{x}_i-\mathbf{x}_j\right\|^2/t\right)$$

Belkin M., Niyogi P., “Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering”, NIPS 14, 2002.

主成分分析 (PCA)



- 主成分分析是一种线性降维技术
  - 在几何上看，是进行坐标系变换
  - 使用样本协方差矩阵的主特征向量作为坐标轴，定义一个新坐标系，然后把数据投影到这个新坐标系中

- 给定数据矩阵X $\mathbf{X}=\left(\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_n\right)$， 其中 $\mathbf{x}_i\in R^m$
  - 计算样本协方差矩阵**S** $$S=\frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{x}_i-\boldsymbol{\mu}\right)\left(\mathbf{x}_i-\boldsymbol{\mu}\right)^T$$
  - 对矩阵**S**进行特征值分解 $S\cdot\mathbf{p}_i=\lambda_i\mathbf{p}_i$， $\mathbf{p}_i\in R^m, i=1,\ldots,r$
  - 取前**r**个最大特征值对应的特征向量定义投影矩阵**P**，低维特征可以获得

$$\mathbf{Y}=P^T X$$

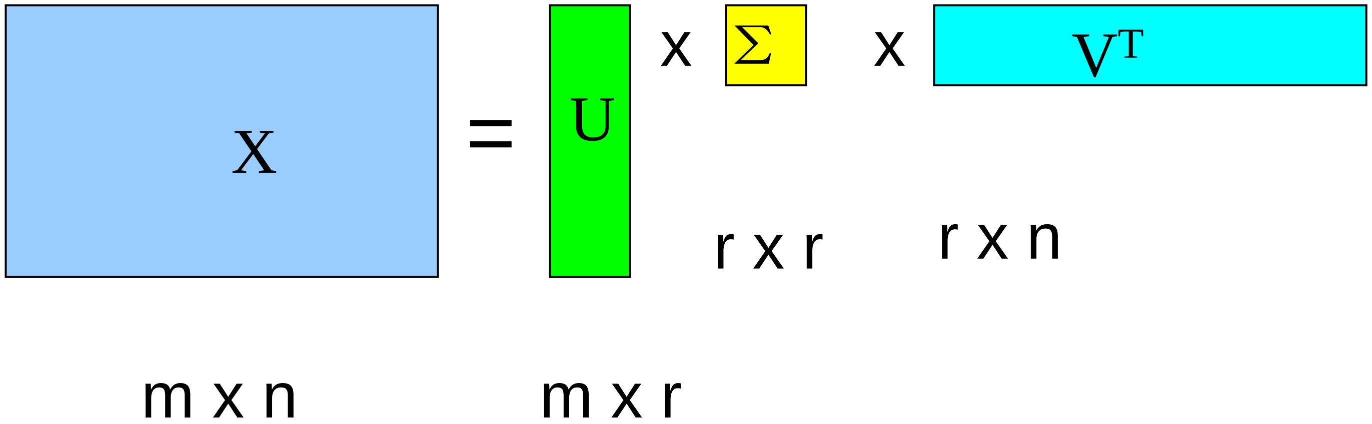
其中 $P=\left(\mathbf{p}_1,\mathbf{p}_2,\cdots,\mathbf{p}_r\right)$， $\mathbf{Y}=\left(\mathbf{y}_1,\mathbf{y}_2,\cdots,\mathbf{y}_n\right)$， $\mathbf{y}_i\in R^r$

矩阵的低秩近似

- 如果矩阵X的秩是r，则可以采用紧凑SVD表示

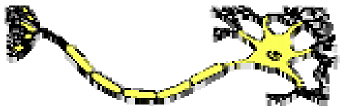
$$X = U_{m\times r}\Sigma_{r\times r}V_{n\times r}^{T}$$

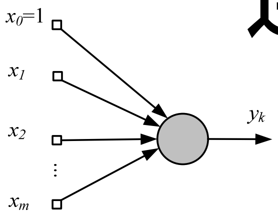
  - **若X的秩大于r，我们仍可使用秩r矩阵去近似矩阵X(其中r远小于m和n)**

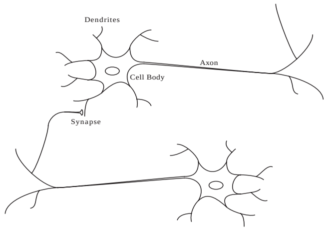


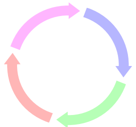
  - **若要压缩空间来表示原矩阵X，只需存储U, Σ 和V**

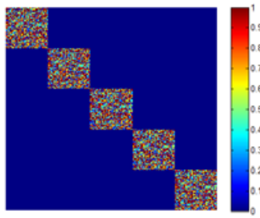
连连看?

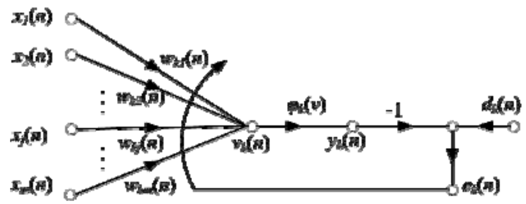


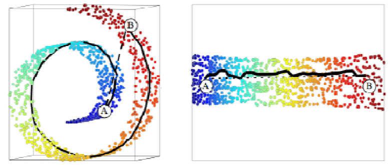

x_0=1
x_1
x_2
x_m
y_k

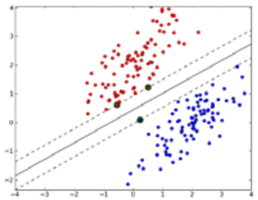

Dendrites
Axon
Cell Body
Synapse

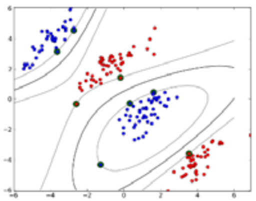


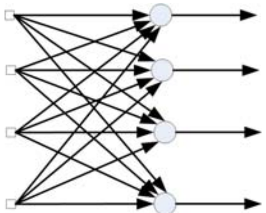


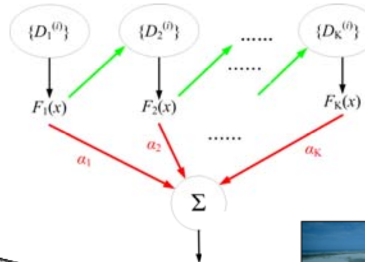


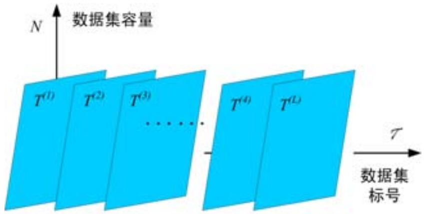


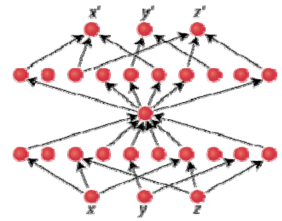


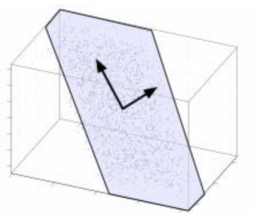


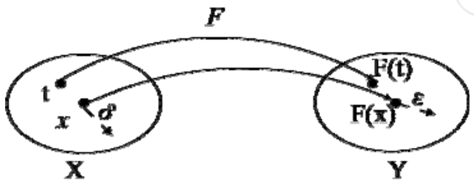


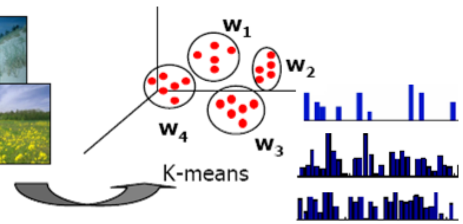

$F_1(x)$
$F_2(x)$
$F_K(x)$
Σ

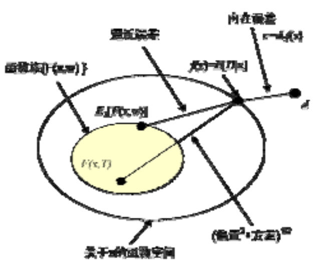

数据集容量
数据集
标号

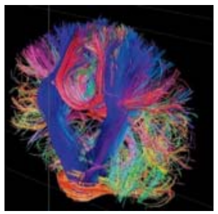


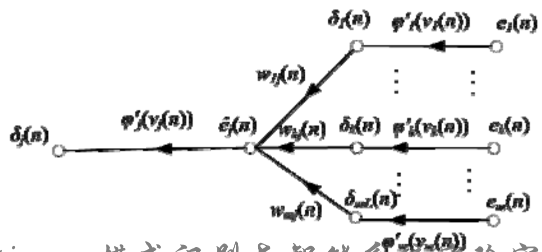


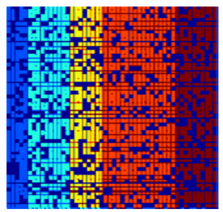

F
F(t)
F(x)
X
Y

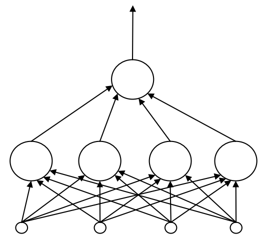

K-means









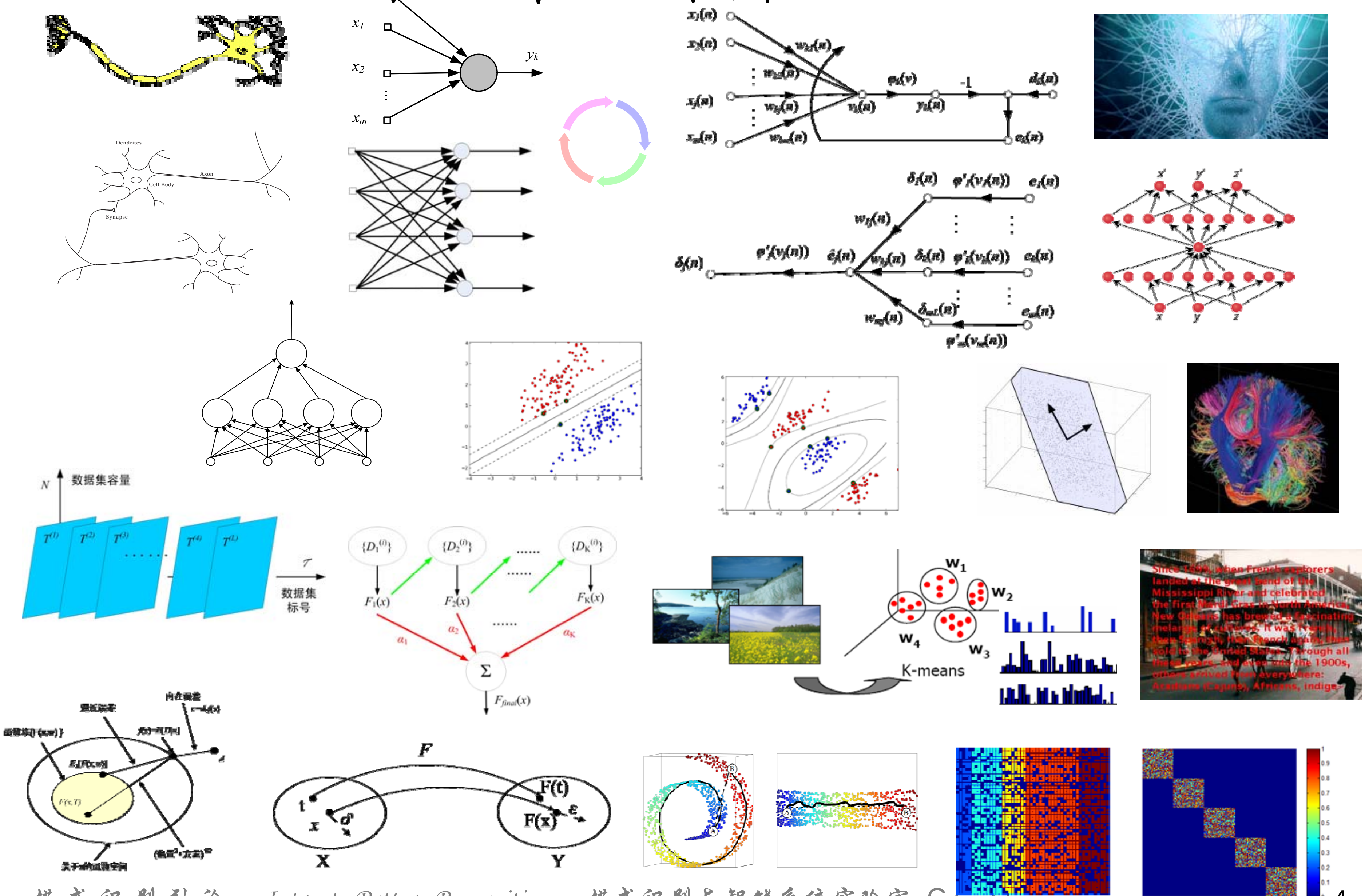


神经计算：学习机器入门

Neural Computation：An Introduction to Learning Machines

主讲: 李春光

图说神经计算 2015



Dendrites
Cell Body
Axon
Synapse
数据集容量
数据集标号
K-means
F
F(t)
F(x)
X
Y

课程结束语

- 统计学习理论与人生轨迹
  - 在学习专业知识的同时，要学习如何做人、做事、做学问。学习专业知识相当于经验误差最小项，做人做事做学问的道理相当于今后人生轨迹的正则化项。
    - 无论将来遇到什么情况，都要对未来保持一个美好的希望；无论将来在什么工作岗位，要保持正直与诚实，要追求一种专业精神。

"No matter what you do, your attitude determines your altitude". (无论做什么事情，你的态度决定你的高度)

• 谢谢大家！