

Recommender System 简介

Yurong Tao

什么是Recommender System

什么是豆瓣猜 ·····

这是豆瓣给你的个人推荐。通过每天分析你的读过、想读、在读、评价行为，豆瓣会从海量数据里挑选你会感兴趣的内容给你。

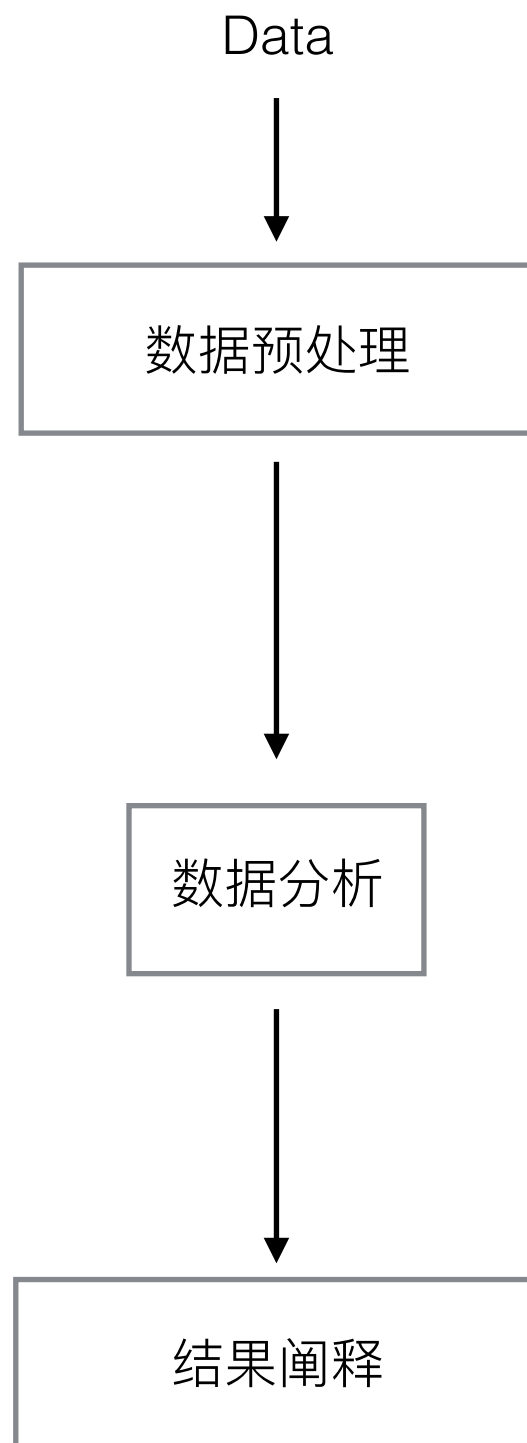
你用得越多，豆瓣猜得就越准确。

推荐系统分类

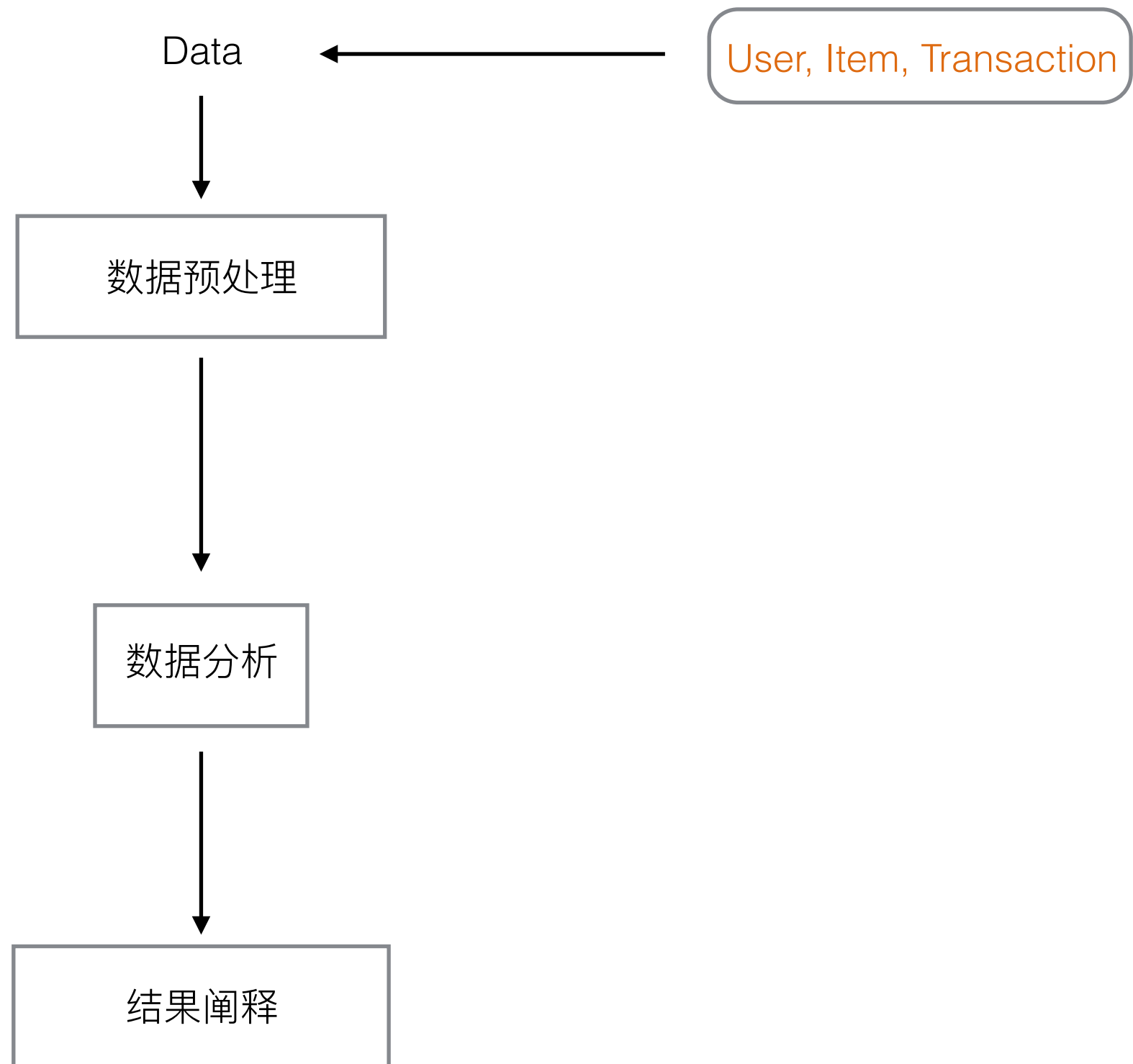
- Content-based – 基于内容的推荐
- Collaborative filtering – 协同过滤推荐系统
- Demographic – 人口统计学的推荐
- Knowledge-based - 基于知识的推荐
- Community-based – 基于社区的推荐系统
- Hybrid – 混合推荐系统



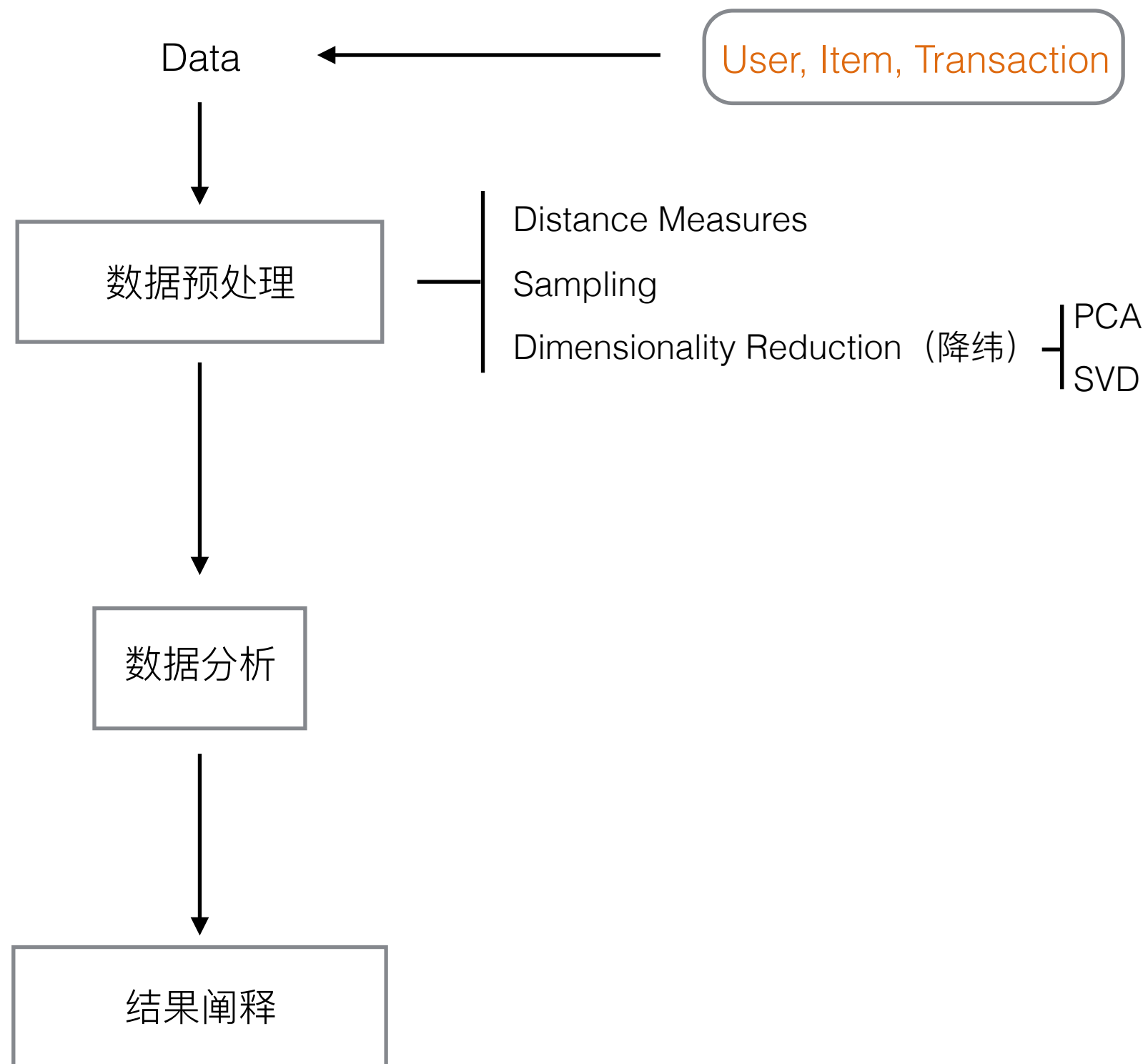
推荐技术



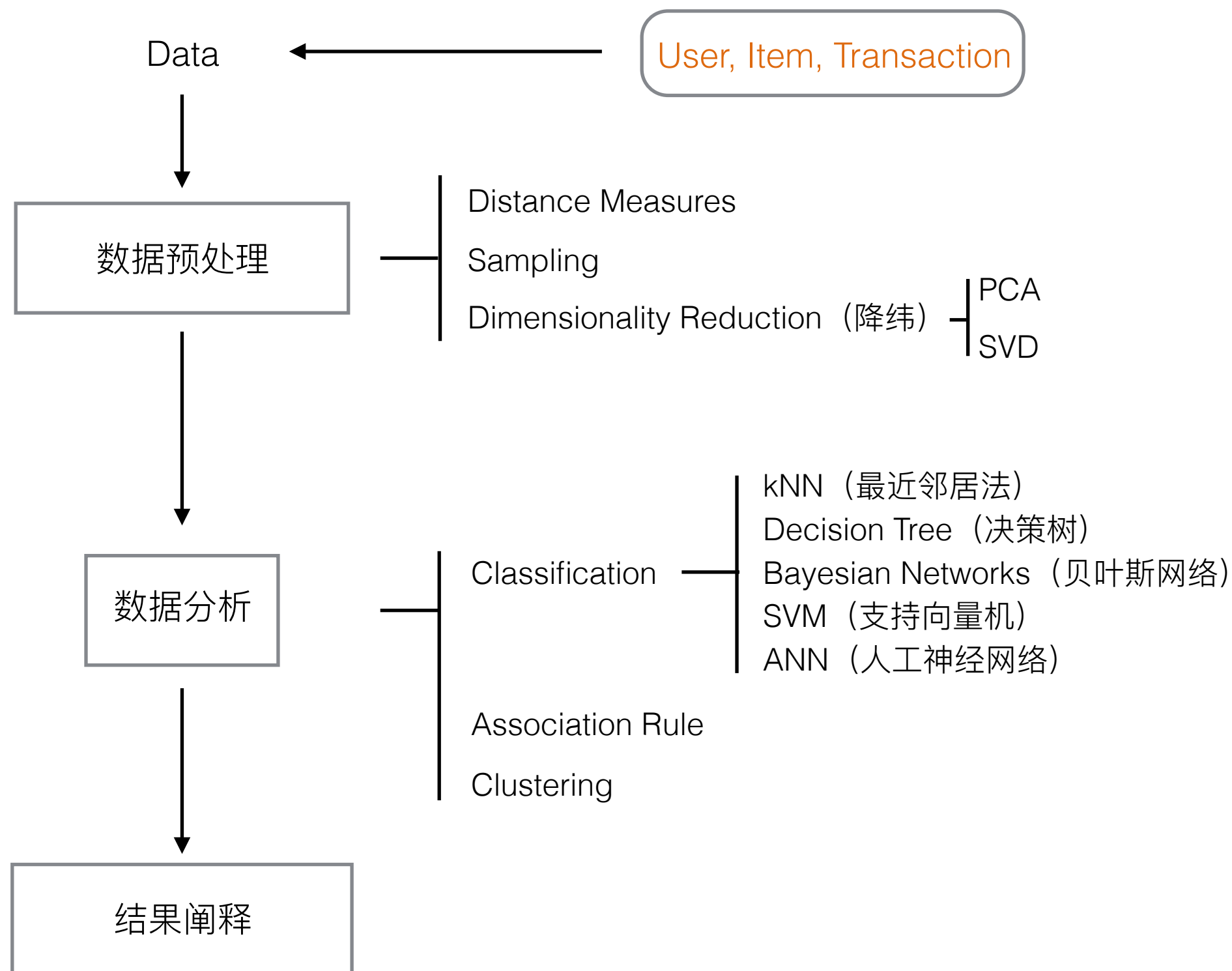
推荐技术



推荐技术



推荐技术



Distance Measures

Euclidean Distance

$$d(x, y) = \sqrt{\sum_{k=1}^n (x_k - y_k)^2}$$

Minkowski Distance

$$d(x, y) = \left(\sum_{k=1}^n |x_k - y_k|^r \right)^{\frac{1}{r}}$$

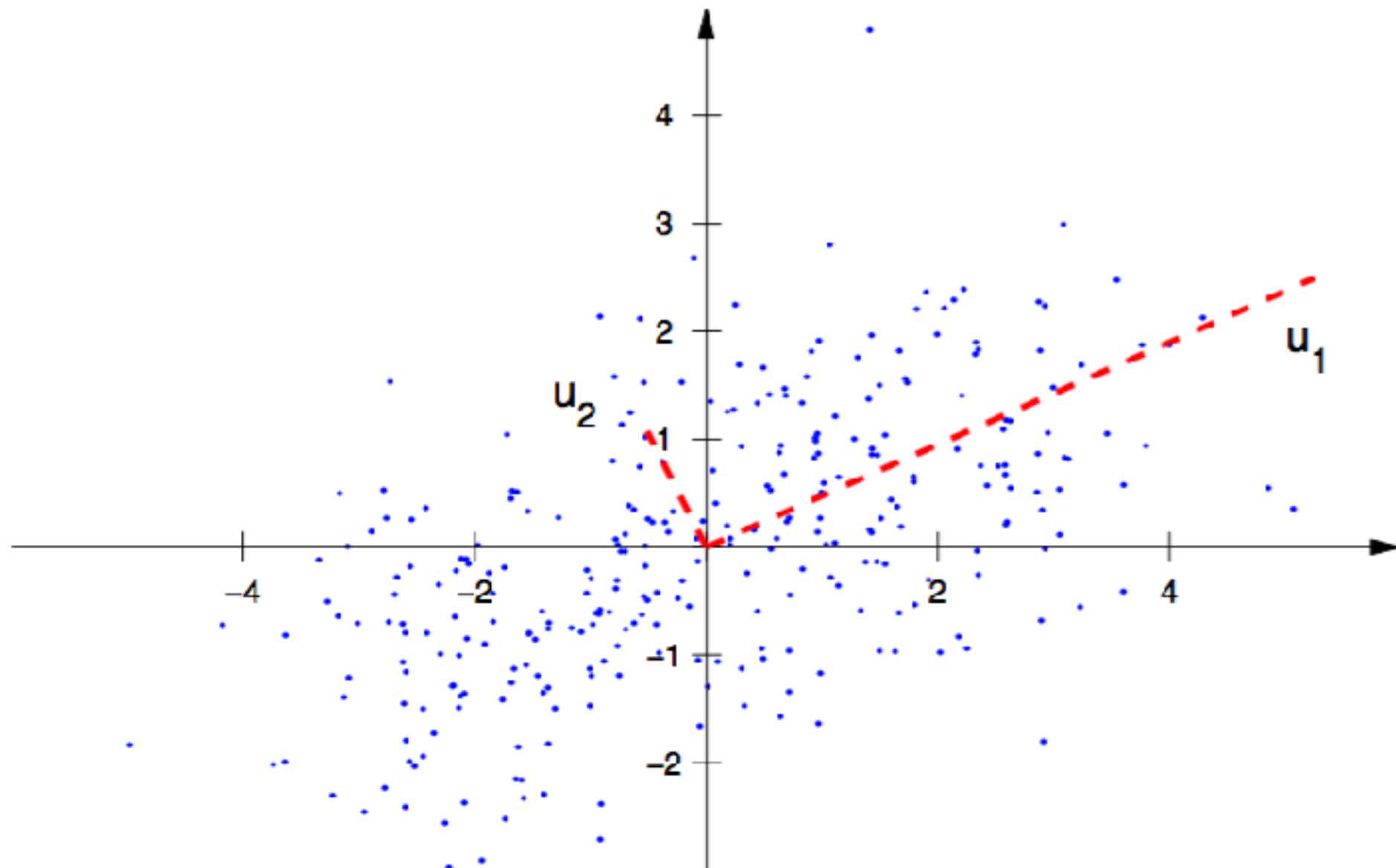
Mahalanobis distance

$$d(x, y) = \sqrt{(x - y) \sigma^{-1} (x - y)^T}$$

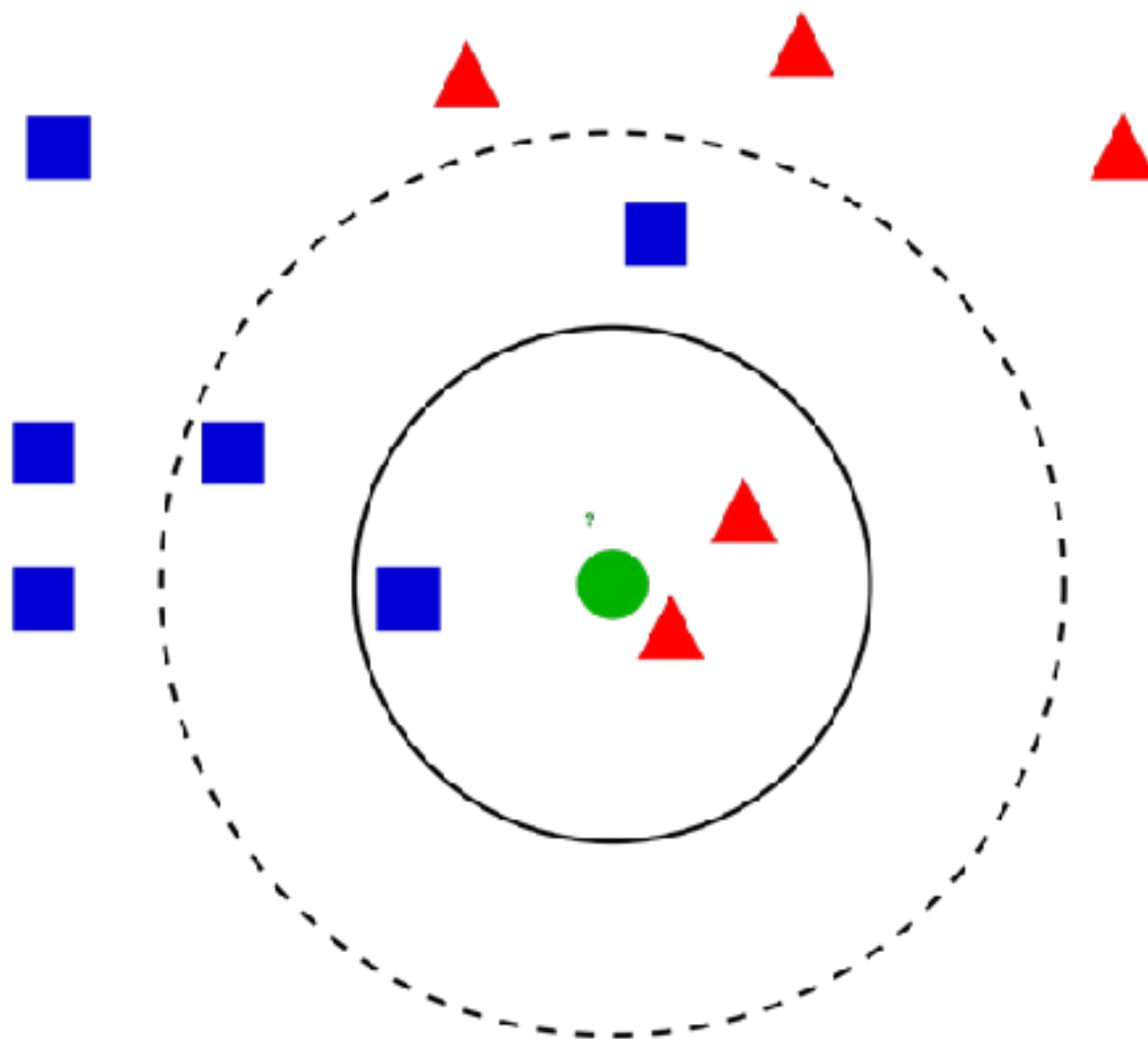
Cosine Similarity

$$\cos(x, y) = \frac{(x \bullet y)}{\|x\| \|y\|}$$

PCA & SVD



kNN (最近邻居法)



Bayesian Networks (贝叶斯网络)

可以用来

- 检测垃圾邮件
- 做文本分析
- 根据身高、体重、鞋码猜测此人的性别

基于贝叶斯理论

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)}$$

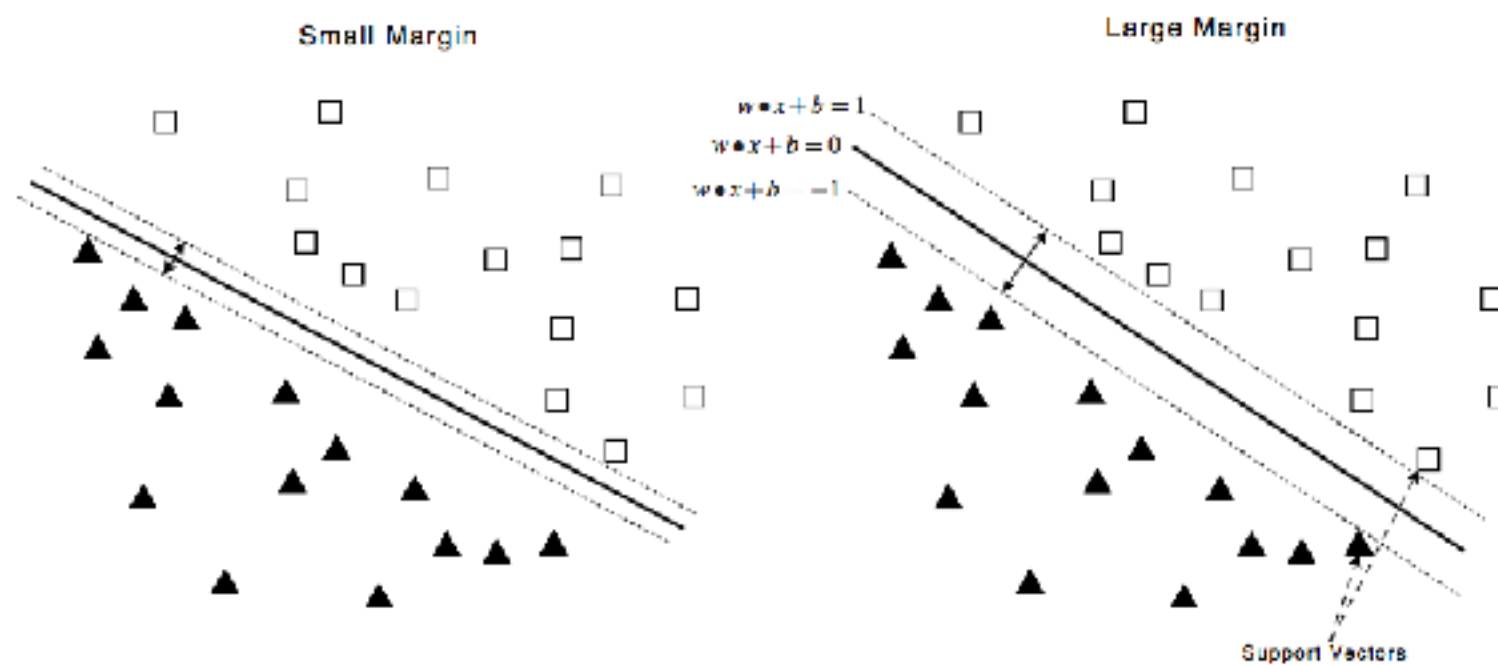


贝叶斯分类器

$$p(C|F_1, \dots, F_n) = \frac{1}{Z}p(C) \prod_{i=1}^n p(F_i|C)$$

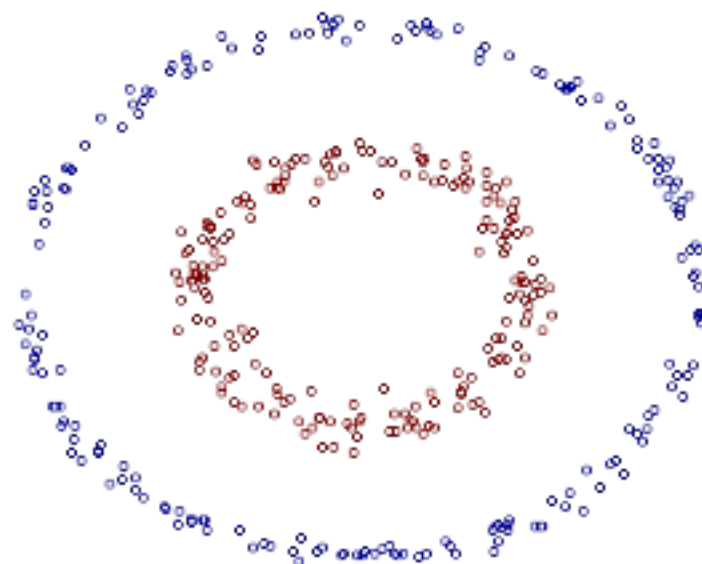
SVM (支持向量机)

线性

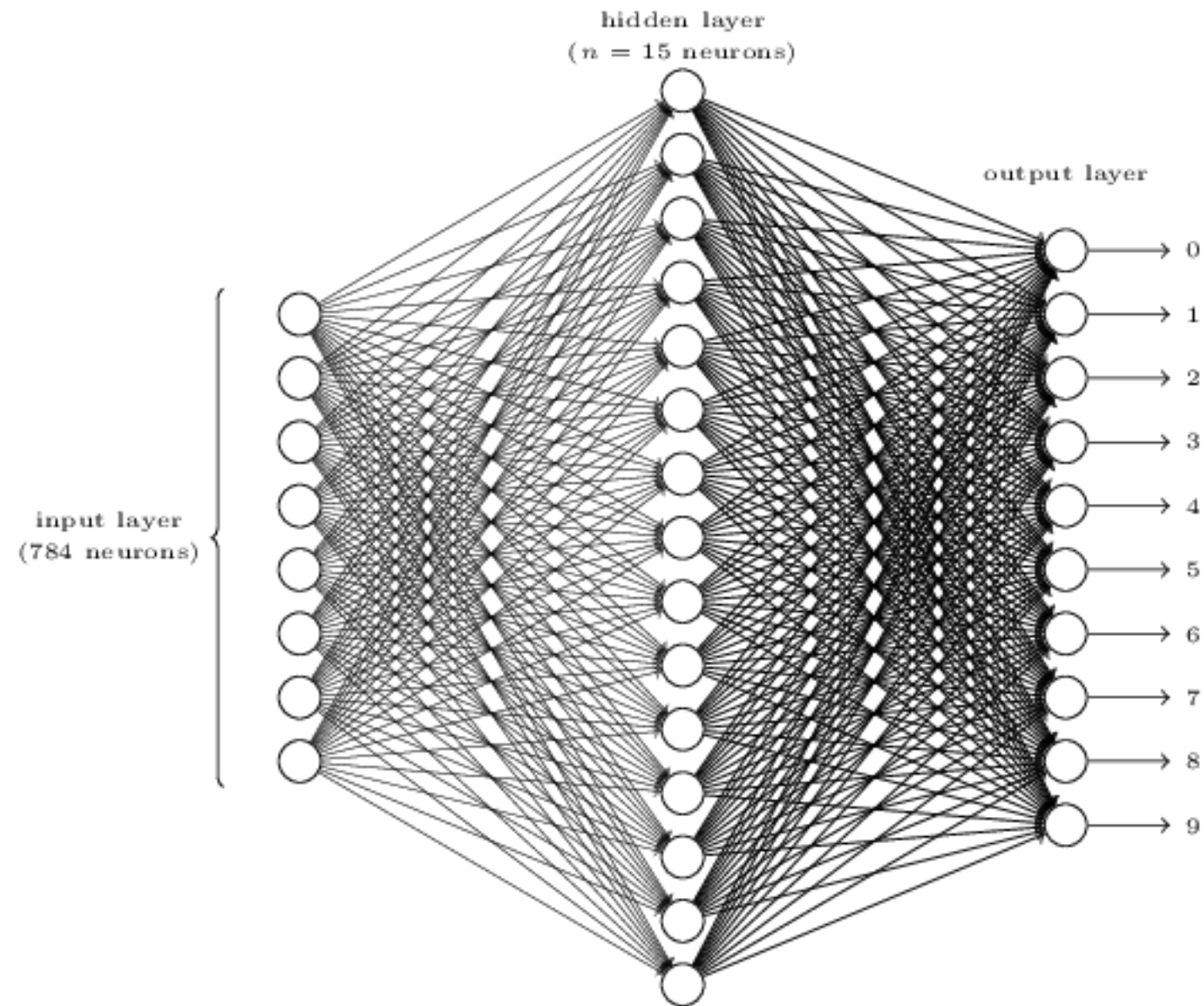


非线性

<https://www.youtube.com/watch?v=3liCbRZPrZA>



ANN (人工神经网络)



Demo