P2:

TB图蓝线表示没100000例中有多少例的TB，红线表示每年TB人数的百分比变化

全球变暖图纵轴是与一个基准年的温度差

语音的纵轴还不太清楚

纽交所图的纵轴是每天大盘的跌涨

P5:

三个不同的划分段的标准

P7:

Linear Regression会导致线段不连续

P11:

Univariate指的是只基于一个时间序列过去和现在值来预测未来值

Multivariate指的是基于多个时间序列的值来预测，比如销量同股票价格和经济指数相关

P12

SES:用过去值的weighted sum来预测未来值，越近期的数据权值越大

递归的形式使得可以根据的预测值和当前值来计算的预测值

:0.1-0.3 用于未来值基于较多历史值的情况

:🡪1 主要用最近期的值来预测未来值

P14-P15不懂略过

P16

Temporal Pattern Cluster的作用是防止pattern定义的太确切，而且pattern也不大可能在每一维的数值上完全相等，图例也可以看出

P17

=1 , Q = 2

P20

P21

t检验用来检验两个样本的总体差异性

optimization不仅要优化还应该有

P22

handwritten documents可以用笔迹坐标的时间序列

query-by-humming哼唱查询

P23

欧式距离实际上L2 distance，在规模较大的时候Lp distance将收敛到欧式距离

P26

=0可以保证Q匹配到X的一个后缀而不是一定要从开始匹配

最终从选一个代价最小的j

P29

Filter Step在计算F(Q)与F(X,j)的距离时，不用计算所有的，而是以L为长度分段，每段计算一个值来作为一个采样

Refine Step只检查X中长度不超过2Q的序列

P30

sequence是事务的序列，顺序至关重要

item set是一个事务中item的集合，顺序无所谓

frequent item set挖掘可以做同一个事务内的关联规则挖掘或推荐，比如你买相机的时候给你推荐镜头一起购买。

frequent sequence挖掘可以做事务外的关联规则挖掘或推荐，即根据历史的交易，推荐下一次交易，比如买完相机一个礼拜，给你推荐各种镜头

P31

参考Apriori算法的思想L1🡪L1 candidate🡪L2🡪L2 candidate🡪…🡪Lk

P40

左边的图在每一个(i,j)位置都计算了一个相关值，右边的图只显示了相关度大于T的

Brute-force需要O(n)次扫描数据库

P42

一次扫描数据库，扫描完毕后把DFT的前5项存在内存中，用来过滤不相关的。

P43

过滤完了之后需要对剩下来的候选对进行验证，需要一个好的读取顺序

P44

中的B代表binary的意思

P45

跟那个item set 和sequence mining一个样看是否在一个事务中

P51

第k个motif必须和前k-1个motif的每一个距离大于2R，否则很有可能两个motif的匹配序列重叠很多，区别不大。

P53

降维划分为w段，每段取均值，n维到w维

离散化去a个值，高斯分布的面积划分为a段，每段去一个symbol

P54

三角不等式

P57

k-gram based，给定一个k-gram set把每个TS表示成k-gram 频繁度的向量

pattern based: 1.至少在一个class里是频繁的

2.至少在一个class里是不频繁的，有特色的

feature mining: 给定pattern的标准挖掘满足标准的feature

P58

SVM的某些kernel可以作为similarity的度量