主旨：Possibilities of (Supervised learning) feature selection

Dimensionality reduction techniques:

Feature selection (不改变原数据，只是选取subset)

Principal component analysis PCA (based on projection)

Information theory (based on compression压缩)

Overviews of feature selection techniques

1, why we need FS?

在图像识别前分类好数据

(a) to avoid overfitting and improve model performance

(b) to provide faster and more cost-effective models

(c) to gain a deeper insight into the underlying processes that generated the data.

2, Instead of just optimizing the parameters of the model for the full feature subset, we now need to find the optimal model parameters for the optimal feature subset,

3，the search in the model hypothesis space is augmented by another dimension: the one of finding the optimal subset of relevant features

FS的分类 (depending on how they combine the FS search with the construction of classification model)

1, filter method 过滤

Univariate 单变量

独立的分析每个特征维度

优：能快速解决高维度数据

缺：忽视了每个维度之间的关系

Multivariate 多变量

2, wrapper method 包装

计算量大

3, embedded method 嵌入

要对项目很了解

应用（4,5点在2007年被称为upcoming domains）

1, Sequence analysis

分类

Content analysis

分类

The prediction of subsequences that code for proteins

提到的算法

Markov models

Interpolated Markov model

Chi-square

Interpolated context model

Bayesian decision tree

Markov blanket multivariate filter approach

The prediction of protein function from sequence

Genetic algorithm & Gamma test

SVM

Recognition of promoter regions

Prediction of microRNA targets

Signal analysis

Binding sites结合位点

Regulatory motifs基序

Translation initiation site翻译初始点

Splice sites剪切位点

Future research: identifying relevant features related to alternative splice sites and alternative TIS

2, microarray analysis (微阵列分析)

基因芯片

3, mass spectra analysis (质谱分析)

4, single nucleotide polymorphism analysis

5, text and literature mining

Dealing with small sample domains

Existing problems

1, small sample size

2, imprecision

3, overfitting

Two ways

1,评价角度

通过选取数据库的一个特征子集，与数据集比较

2,建模角度

Model combination （基于通常有多个有效特征子集）

Feature selection:

<https://blog.csdn.net/hren_ron/article/details/80914491>

<https://www.cnblogs.com/yuesi/articles/9236796.html>

<https://zhuanlan.zhihu.com/p/30404850>

<https://zhuanlan.zhihu.com/p/24635014>