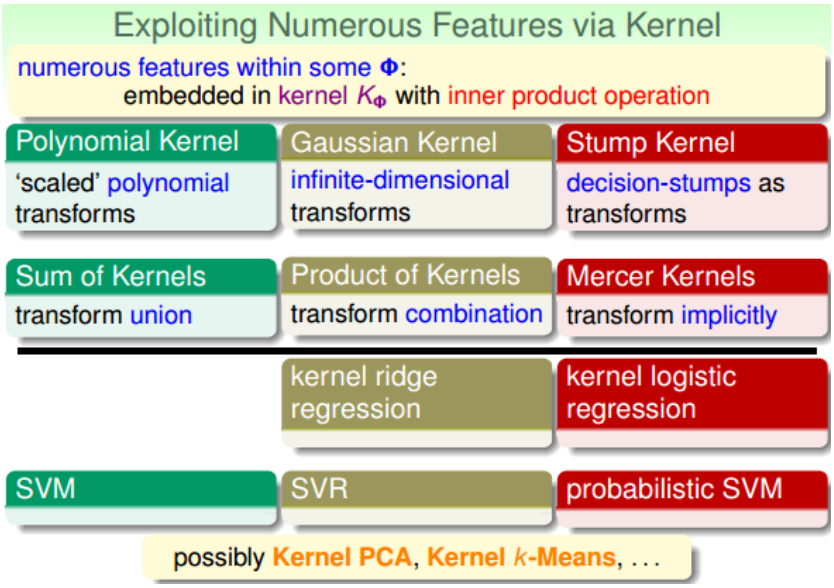
Conclusion

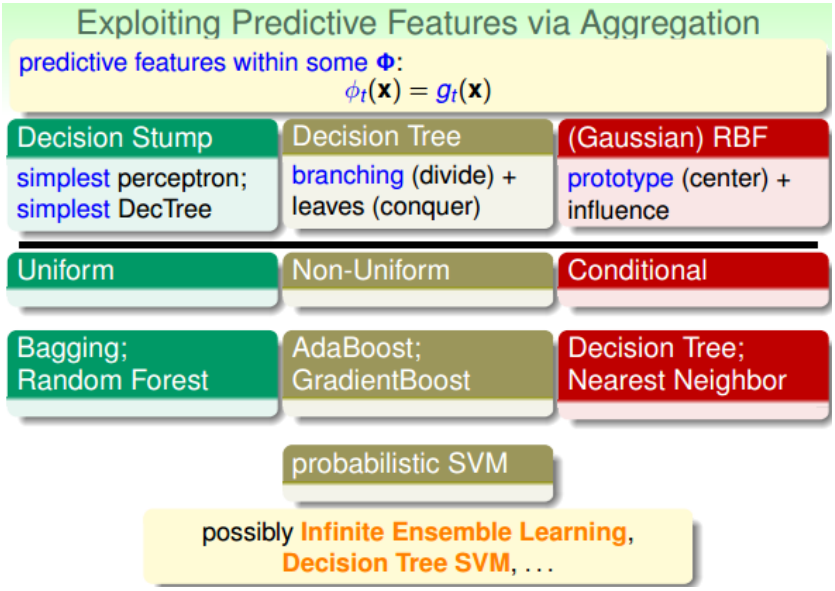
1.Feature Exploitation Techniques

Kernel运算将特征转换和计算内积这两个步骤合二为一，提高了计算效率。
介绍过的kernel有：Polynormial Kernel、Gaussian Kernel、Stump Kernel等。
另外，可以将不同的kernels相加（transform union）或者相乘（transform combination），得到不同的kernels的结合形式，让模型更加复杂。

要成为kernel，必须满足Mercer Condition。
不同的kernel可以搭配不同的kernel模型，比如：SVM、SVR和probabilistic SVM等，还包括一些不太常用的模型：kernel ridge regression、kernel logistic regression。
使用这些kernel模型就可以将线性模型扩展到非线性模型，kernel就是实现一种特征转换，从而能够处理非常复杂的非线性模型。
因为PCA、k-Means等算法都包含了内积运算，所以它们都对应有相应的kernel版本。

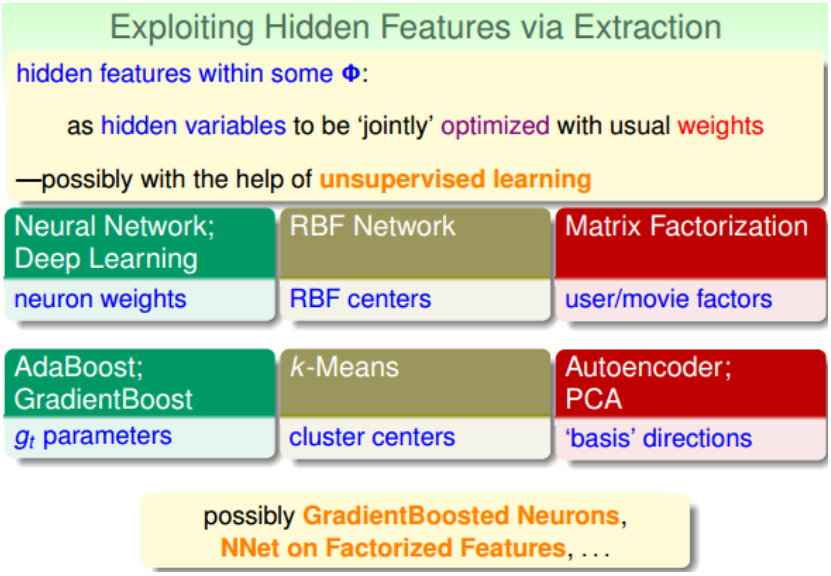


Kernel是利用特征转换的第一种方法，那利用特征转换的第二种方法就是Aggregation。
所有的hypothesis都可以看成是一种特征转换，然后再由这些g组合成G。
分类模型（hypothesis）包括：Decision Stump、Decision Tree和Gaussian RBF等。如果所有的g是已知的，就可以进行blending，例如Uniform、Non-Uniform和Conditional等方式进行aggregation。如果所有的g是未知的，可以使用例如Bagging、AdaBoost和Decision Tree的方法来建立模型。除此之外，还有probabilistic SVM模型。



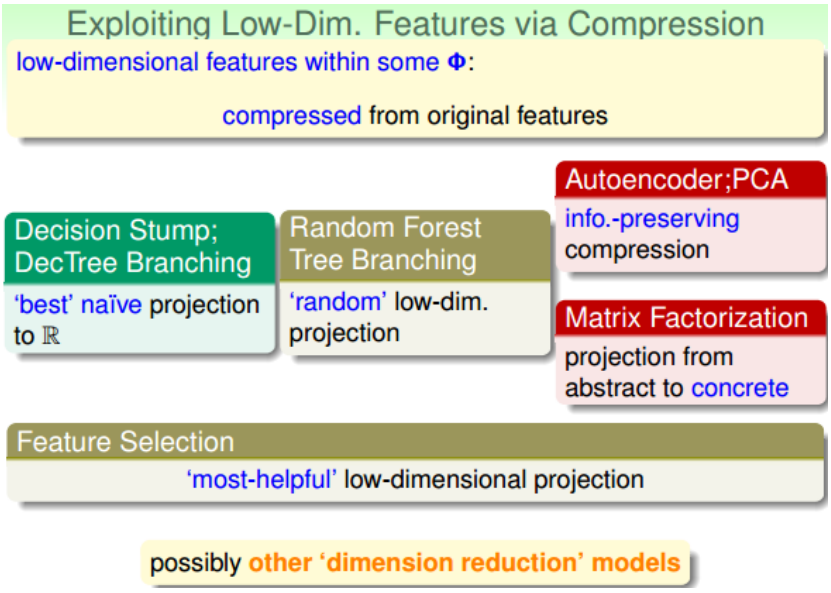
除此之外，我们还介绍了利用提取的方式，找出潜藏的特征（Hidden Features）。

一般通过unsupervised learning的方法，从原始数据中提取出隐藏特征，使用权重表征。
相应的模型包括：Neural Network、RBF Network、Matrix Factorization等。
这些模型使用的unsupervised learning方法包括：AdaBoost、k-Means和Autoencoder、PCA等。



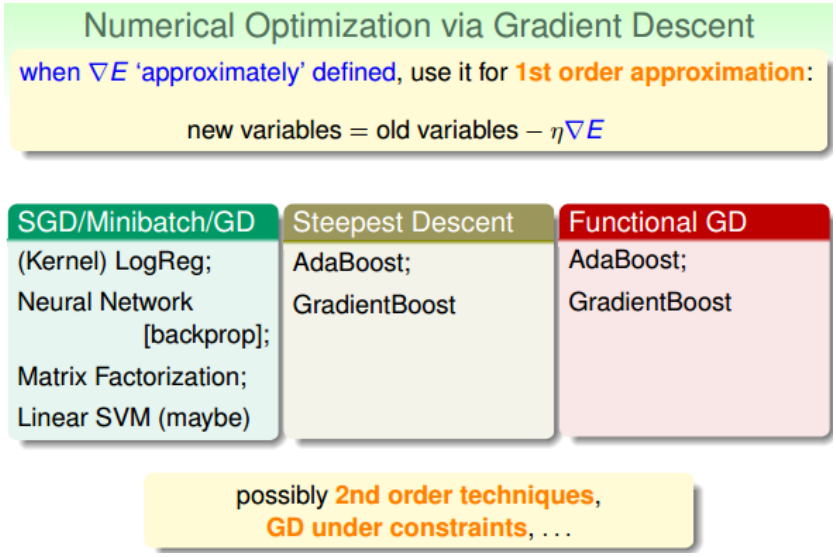
另外，还有一种非常有用的特征转换方法是维度压缩，即将高维度的数据降低（投影）到低维度的数据。

维度压缩模型包括：Decision Stump、Random Forest Tree Branching、Autoencoder、PCA和Matrix Factorization等。
这些从高纬度到低纬度的特征转换在实际应用中作用很大。

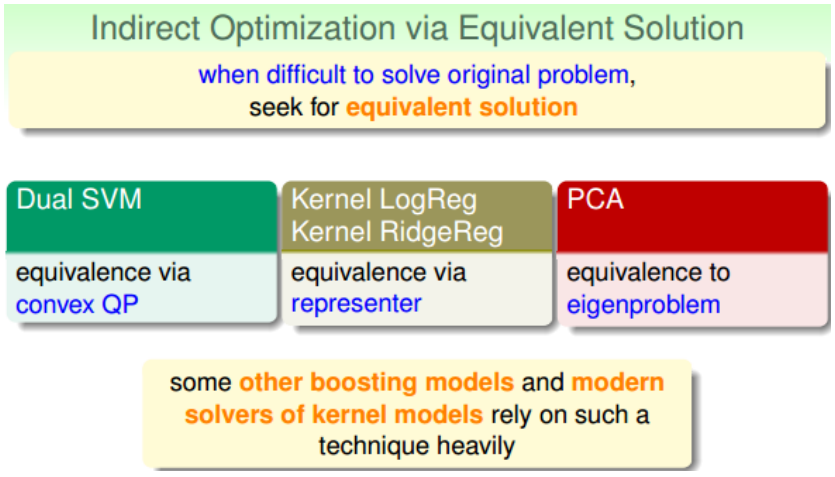


2.Error Optimization Techniques

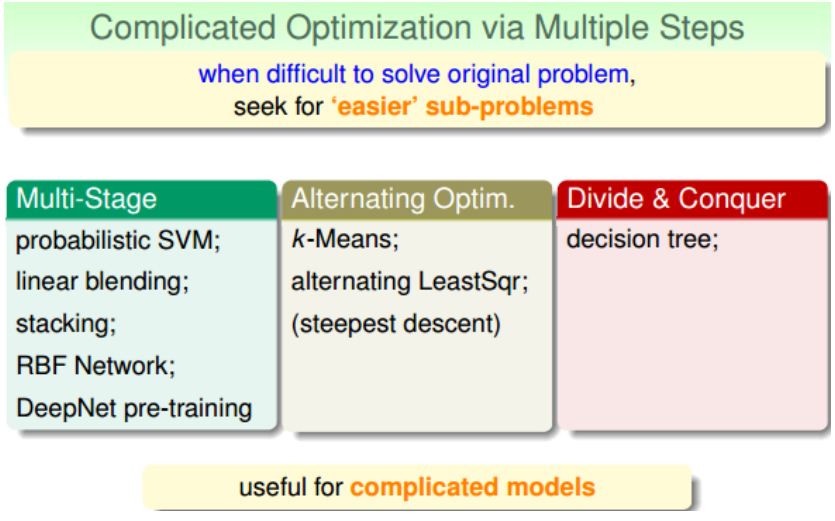
首先，第一个数值优化技巧就是梯度下降（Gradient Descent），即让变量沿着其梯度反方向变化，不断接近最优解。
例如SGD、Steepest Descent和Functional GD都是利用了梯度下降的技巧。



而对于一些更复杂的最佳化问题，无法直接利用梯度下降方法来做，往往需要一些数学上的推导来得到最优解。
最典型的例子是Dual SVM，还包括Kernel LogReg、Kernel RidgeReg和PCA等等。这些模型本身包含了很多数学上的一些知识，例如线性代数等等。
除此之外，还有一些boosting和kernel模型，都会用到类似的数学推导和转换技巧。



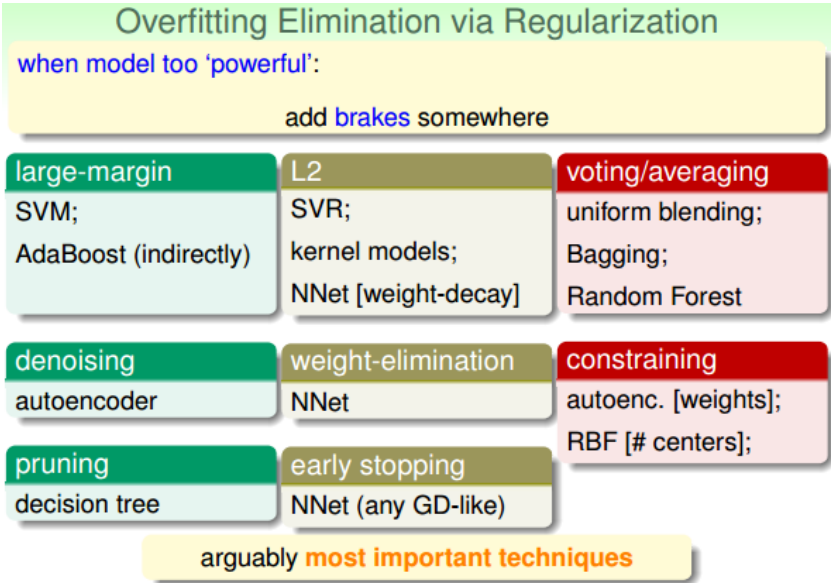
如果原始问题比较复杂，求解比较困难，可以将原始问题拆分为子问题以简化计算。也就是将问题划分为多个步骤进行求解，即Multi-Stage。
例如probabilistic SVM、linear blending、RBF Network等。
还可以使用交叉迭代优化的方法，即Alternating Optim。例如k-Means、alternating LeastSqr等。
除此之外，还可以采样分而治之的方法，即Divide & Conquer。例如decision tree。



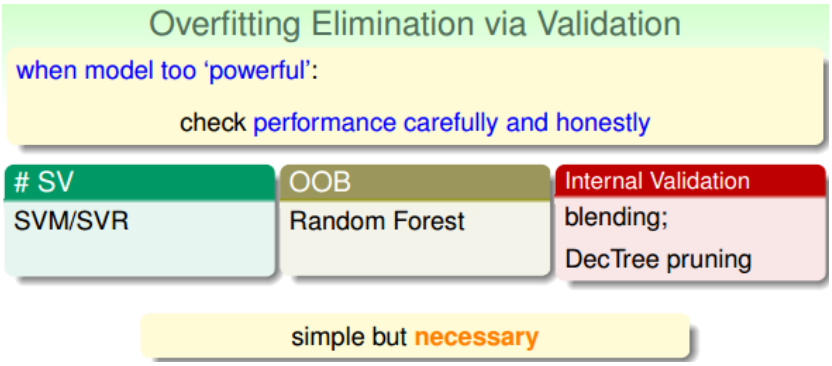
3.Overfitting Elimination Techniques

Feature Exploitation Techniques和Error Optimization Techniques都是为了优化复杂模型，减小E_{in}。但是E_{in}太小有很可能会造成过拟合overfitting。因此，机器学习中，Overfitting Elimination尤为重要。

首先，可以使用Regularization来避免过拟合现象发生。方法包括：large-margin、L2、voting/averaging等等。



除了Regularization之外，还可以使用Validation来消除Overfitting。
Validation包括：SV、OOB和Internal Validation等。



4.Machine Learning in Action

本小节介绍了台大团队在近几年的KDDCup国际竞赛上的表现和使用的各种机器算法。

NTU KDDCup 2010 World Champion Model

Feature engineering and classifier ensemble for KDD Cup 2010, Yu et al., KDDCup 2010

linear blending of

Logistic Regression + many rawly encoded features

Random Forest + human-designed features

yes, you've learned everything! :-)

NTU KDDCup 2011 Track 1 World Champion Model

A linear ensemble of individual and blended models for music rating prediction, Chen et al., KDDCup 2011

NNet, DecTree-like, and then linear blending of

- Matrix Factorization variants, including probabilistic PCA
- Restricted Boltzmann Machines: an 'extended' autoencoder
- *k* Nearest Neighbors
- Probabilistic Latent Semantic Analysis: an extraction model that has 'soft clusters' as hidden variables
- linear regression, NNet, & GBDT

yes, you can 'easily' understand everything! :-)

NTU KDDCup 2012 Track 2 World Champion Model

A two-stage ensemble of diverse models for advertisement ranking in KDD Cup 2012, Wu et al., KDDCup 2012

NNet, GBDT-like, and then linear blending of

- Linear Regression variants, including linear SVR
- Logistic Regression variants
- Matrix Factorization variants
- ...

'key' is to **blend properly without overfitting**

NTU KDDCup 2013 Track 1 World Champion Model

Combination of feature engineering and ranking models for paper-author identification in KDD Cup 2013, Li et al., KDDCup 2013

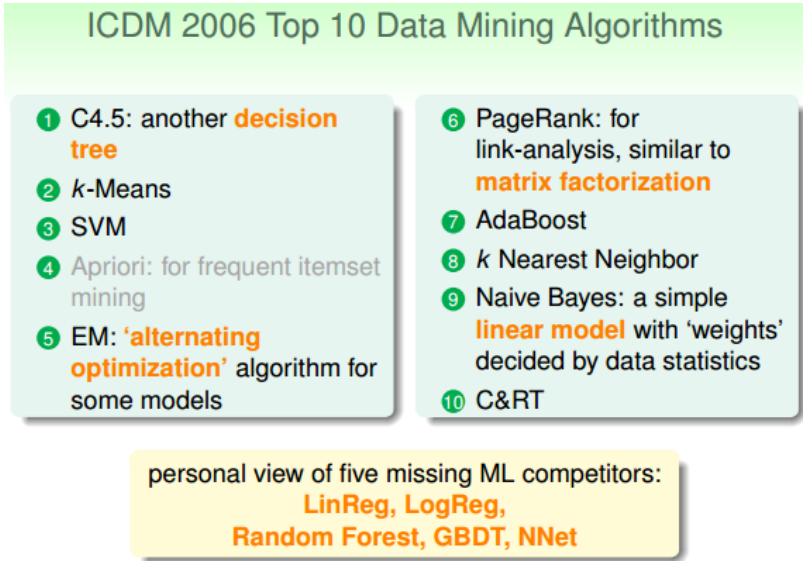
linear blending of

- Random Forest with many many many trees
- GBDT variants

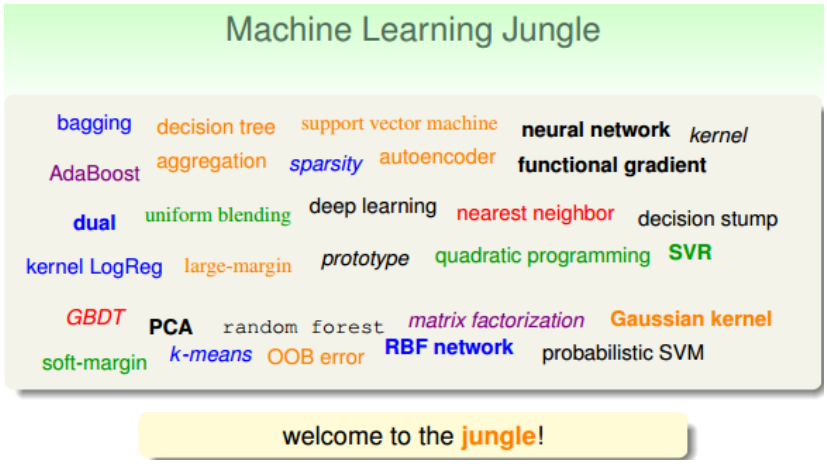
with tons of efforts in designing features

'another key' is to **construct features with domain knowledge**

ICDM在2006年的时候发布了排名前十的数据挖掘算法



最后，将所有介绍过的机器学习算法和模型列举出来：



5.Summary