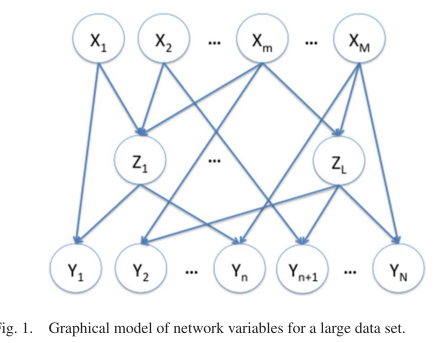
传统的数据分析并没有关注数据产生的机制以及数据之间的联系，这篇文章将通信理论和信息论加入其中，考虑数据的生成。  
  
换一种想法，机器学习中有要素X,Y，不妨将X和Y中信息的传递想象成一个经典的点对点通信问题



该图显示了一个大的数据集中变量的图模型，然而这个图往往是不稳定的，所以这张图应该是一个概率图】

概率图模型可以使用机器学习的思想来处理，也可以看作是通信的结构。

将X1,X2,…..XN和Y之间的因果关系问题看作是多址通信中的信息耦合问题。

问题建模：

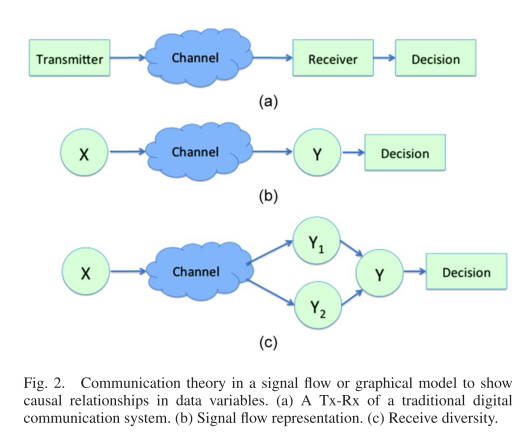
Problem: To infer Y based on X1,X2,…,XN

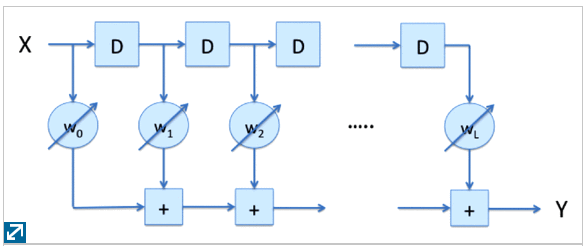
1: 使用均衡器来解决变量之间的关系。X1到Y,X2到y 等等

2: 根据最小均方误差选中额变量个数，并在线更新相关性矩阵

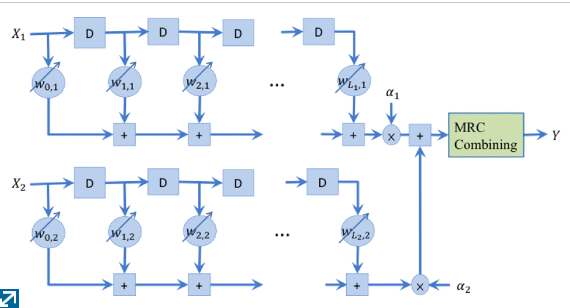
3: 将各个独立的系统连接起来。

这个模型关注点在优化信息传送以及避免数据变量之间的交叉干扰。，现有的统计学习方式将所有的数据变量都加以利用，这种方式将引起多址干扰。也就是说，系统性能的好坏和使用变量的多少并不成正比（干扰可以看作机器学习中的过拟合）





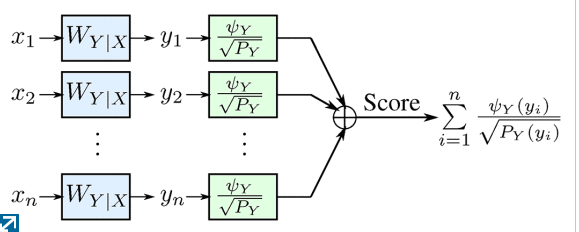
使用均衡器的方式来对连续数据进行预测，该图仅仅是一个简单的线性均衡器、



使用上述结构对多个特征进行联合预测

信息耦合：

C:\Users\TinyLee\AppData\Local\Temp\1489460449(1).png 这个公式很棒



图中绿框表示的是对Y的降维统计表示,他们的和可以很好的表示信息

其中 is the singular vector of DTM with the largest singular value

数据降维