



波现象与智能反演成像研究组



学习类算法的本质问题思考

从数据处理拓展到信息处理

报告人：王华忠

波现象与智能反演成像研究组 (WPI)

同济大学海洋与地球科学学院，上海

2021年1月5日

目录

- ◆一、概述
- ◆二、Learning From data的数学基础
- ◆三、非学习类算法及本质特征分析
- ◆四、学习类算法及本质特征分析
- ◆五、关于学习类算法、ML及AI的思考
- ◆六、勘探地震数据分析中的ML算法探讨
- ◆七、总结与讨论



◆一、概述

- ◆学习类算法或多或少是模仿人类或生物智能的。
- ◆人类的大脑，至少在当前人类所感知到的宇宙范围内的生物中，学习能力是最强的。
- ◆人类梦寐以求想回答如下问题：
 - ◆人脑是如何工作的？
 - ◆人类如何从现实世界获取知识和运用知识？
 - ◆我们能否制作出模仿人脑的计算机或计算机算法？



◆一、概述

- ◆**人工神经网络是当前我们对大脑学习能力的进行模仿的一种尝试。**
- ◆**人工神经网络的研究内容：**
 - ◆**生物原型研究；**
 - ◆**建立生物原型的、抽象的理论模型的问题研究；**
 - ◆**网络构型及学习算法（或训练算法）的研究；**
 - ◆**人工神经网络的应用研究。**



◆一、概述

◆人脑擅长的是概念思维（基于模式的思维），“人脑是一个信息处理机”；Von Nuemann构型的计算机擅长的是数据分析（或 Learning from Data），计算机是一个数据处理机。

◆让计算机模拟人的某种学习和决策能力是当今大数据时代的必然要求。

◆发展能融合人的某种学习和决策能力，以及计算机强大的数据分析能力的新构型的计算机是必然的需求。

◆因此，对如何制造AI计算机的探索一直会热度不减。



◆一、概述

- ◆我认为也可以在Von Nuemann构型的计算机上用算法来实现AI计算机的功能。
- ◆这应该就是AI/ML/DL的理想目标。
- ◆不过，更现实的选择应该是在Von Nuemann构型的计算机上发展具有一定学习能力的算法。更具体地，在大数据分析算法的基础上，发展出能在一定程度上模拟人脑决策能力的计算机算法。



◆一、概述

◆In God we trust, all others bring data.

- ◆ William Edwards Deming (1900-1993)
- ◆ 戴明博士的生平：威廉·爱德华·戴明 --- 1900年10月14日生于美国爱荷华州的苏市(Sioux City, IA)，病逝于1993年12月20日，一生岁月几乎涵括了整个20世纪。戴明出生不久，他的父亲威廉得到了怀俄明州新开放的自由土地之后，举家迁往荷第市，然后再到鲍威尔市。少时的Deming家算是贫穷，因此他在少年时代可说是一直在打工，他在12岁的时候有了第一份职业。有时候在外面点亮街灯、除雪，赚起每天一块两毛的工资或在饭店内打杂、洗床每小时工资美金两毛伍等以补家计。
- ◆ 戴明颇负正义感，曾经参加墨西哥边境一个小战争的志愿兵且已搭车赶赴战场，但是后来被发现只有十四岁，因不符规定才被遣返。幸亏是这样，否则若在战场牺牲，世界上就少了一位伟大的质量管理大师了。戴明博士不但有正义感且深具爱心，在日本指导期间因愤慨一位未善待精神病患的医院院长，而利用他对当时驻日美军的影响力而将此人解聘。
- ◆ 戴明博士于1921年从怀俄明大学(Univ. of Wyoming)毕业后继续前往科罗拉多大学进修，并于1925年修得数学与物理硕士，最后于1928年取得耶鲁大学的物理博士学位。戴明博士在学期间曾经于芝加哥的西电公司霍桑工厂工作时得知当时在贝尔研究所的W. A. 休哈特博士，并于1927年见面后就成为亦师亦友的莫逆之交。
- ◆ 戴明博士毕业后婉拒西电公司的工作机会而应聘到华盛顿的美国农业部的固氮研究所工作。他也曾经利用一年的休假到伦敦大学与R. A. Fisher做有关统计方面的研究。
- ◆ 戴明博士1950年应聘去日本讲学，并将其报酬捐出，而后几乎每年都赴日继续指导，奠定了日本企业界良好的质量管理基础。



◆一、概述

◆我认为：当今时代，由非学习类算法向学习类算法的过渡看来势所难免！

◆首先要思考并回答如下问题：

◆由数据驱动发展出的学习类算法的本质是什么？

◆由数据驱动发展出的学习类算法真正的发展方向是什么？

◆如何定位与评价AI、ML、DL的作用？

◆WPI的研究如何由非学习类的算法逐渐转移到二者并重，最后逐渐发展到以学习类算法为主？

目录

- ◆一、概述
- ◆二、Learning From data的数学基础
- ◆三、非学习类算法及本质特征分析
- ◆四、学习类算法及本质特征分析
- ◆五、关于学习类算法、ML及AI的思考
- ◆六、勘探地震数据分析中的ML算法探讨
- ◆七、总结与讨论



◆二、 Learning From data的数学基础

◆概率统计学---从数据中学习的基本数学框架

- ◆概率统计学的产生目的就是建立起Learning from Data的理论框架！
- ◆概率统计学的研究对象是随机过程（随机变量）。
- ◆实际问题中，我们只能获得随机过程（随机变量）的若干次（甚至一次）具体实现（统计学上称一次采样！）。基于这样的小样本，试图理解或把握数据中蕴含的信息。
- ◆理论上，对随机数据的大样本采样可以完整地把握住随机过程（随机变量）所蕴含的统计信息。
- ◆很遗憾，随机数据（随机过程）的采样是昂贵的。一般地，我们仅有小样本。



◆二、 Learning From data的数学基础

◆概率统计学---从数据中学习的基本数学框架

◆实测数据被视为随机过程（随机变量）。随机过程的联合概率密度函数是描述数据蕴含的规律性的最基本数学模型。

◆但是，联合概率密度函数应用起来并不方便、不直观。

◆针对实测的离散数据，从数据中寻找有用信息的基本方法是估计或近似计算实测的随机数据中的各阶统计量。一般地，各阶统计量就是数据中蕴含的信息！

◆统计量往往是（联合）概率密度函数的参变量（潜变量）。

◆实测数据是随机过程的采样样本或具体实现。

◆大样本包含了全部的统计特征。但是实测数据总是小样本！



◆二、 Learning From data的数学基础

◆概率统计学---从数据中学习的基本数学框架

- ◆因此，从数据中获取有用信息的最根本理论基础，就是概率统计学！
- ◆在大数据分析时代，对概率统计学的学习和理解是必须要强化的！



◆二、 Learning From data的数学基础

◆Gareth James • Daniela Witten • Trevor Hastie • Robert Tibshirani

◆An Introduction to Statistical Learning with Applications in R

◆Statistical Learning

◆Linear Regression

◆Classification

◆Resampling Methods

◆Linear Model selection and regularization

◆Moving beyond linearity

◆Tree-Based methods

◆Support Vector Machines

◆Unsupervised Learning

统计学习的内容似乎与ML是一样。但统计学习更强调概率统计理论。ML更强调优化算法。

目录

- ◆一、概述
- ◆二、Learning From data的数学基础
- ◆三、非学习类算法及本质特征分析
- ◆四、学习类算法及本质特征分析
- ◆五、关于学习类算法、ML及AI的思考
- ◆六、勘探地震数据分析中的ML算法探讨
- ◆七、总结与讨论



◆三、非学习类算法及本质特征分析

- ◆从概率统计的角度基于观测数据来归纳总结规律，必须有大量样本，而样本往往很难全面地获取，而基于小样本的统计，一般地，得不到正确的规律。
- ◆为了克服大样本才能得到统计规律的问题，有必要引入对随机信号的先验假设。首先，认为实际观测的数据由确定性信号加噪音而形成： $\mathbf{d}^{obs} = \mathbf{s} + \boldsymbol{\eta}$,
- ◆假定 \mathbf{s} 是可以由某种已知的（或某种人为建立的）模型来预测的，假定噪音 $\boldsymbol{\eta}$ 满足某种既定的概率分布的。
- ◆针对信号 \mathbf{s} 的模型建立过程，也可视为从数据获得信息或规律的过程。但这与前面讲到的从数据中得到统计规律是不同的。
 - ◆所建立的关系，不能认为一定是因果关系！ 仅仅还是某种预测关系！



◆三、非学习类算法及本质特征分析

◆非学习类的方法往往假设已经知道了表达信号的模型，只是模型中包含的系数不知道。譬如对于由物理系统产生的信号，描述物理系统的状态响应的数学模型往往是偏微分方程或积分方程。而对于信号/图像分析问题，模型一般由基函数叠加形成的。多项式函数应该也是广泛可用的基函数，譬如样条函数广泛用于图像分析问题的建模中。

◆非学习类算法的三种典型反问题模型：

◆图像处理中ROF模型： $J(u) = \frac{1}{2}(Au - g)^2 + \alpha TV(u)$ 。对于图像去噪问题， $A=I$ （单位矩阵）。

◆信号处理中去噪或反褶积模型： $J(\beta) = \frac{1}{2}(\Psi\beta - d^{obs})^2 + \alpha \|\beta\|_1$

◆物理系统参数估计模型： $J(m) = \frac{1}{2}(A(m) - d^{obs})^2 + \alpha \Omega(m)$ 。 $\Omega(m)$ 代表不同形式的正则化方式。

◆在Bayes估计理论下，把微分方程/积分方程中的（物理）参数估计出来，与把基函数叠加模型中的系数估计出来，算法本质上是没有任何差别的。

◆不过，一般地，我们对（物理系统）参数会有更多的先验认识，能提出更多的、基于物理的正则化约束。而对于基函数叠加系数，基本上仅能施加稀疏性约束。这是信号分析、图像分析与物理系统参数估计很不一样的地方。



◆三、非学习类算法及本质特征分析

◆非学习类算法的一个典型特征是：

◆没有泛化应用能力，不同的观测数据，算法要重复执行。

◆可以看出，非学习类算法的规律性、统一性是很强的。核心就是求解一个反问题：在Bayes估计理论下，求解物理数学方程中包含的系统参数的估计问题，或求解对信号或图像的最佳建模问题（基函数选择和叠加系数最佳估计问题）。

◆目前，正在非线性、非高斯、非平稳假设的方向上不断深化，但发展势头已经不太强劲。

◆目前研究和应用的重点，似乎正在向学习类算法转移。

◆但是，我认为WPI当前还是要在非学习类方向上深入下去，这是我们当前的立脚点。

目录

- ◆一、概述
- ◆二、Learning From data的数学基础
- ◆三、非学习类算法及本质特征分析
- ◆四、学习类算法及本质特征分析
- ◆五、关于学习类算法、ML及AI的思考
- ◆六、勘探地震数据分析中的ML算法探讨
- ◆七、总结与讨论



◆三、学习类算法及本质特征分析

◆学习类算法到底有哪些独特的地方？

◆我个人的观点：

◆a) 学习类算法随数据样本的训练不断进化，体现出类似人类学习的某些特征。

◆b) 学习类算法的构建，目的是应用于判决（决策），而不是非学习类算法的参数估计。

◆据此可以看出：学习类算法和非学习类算法还是存在重大差异的。

◆目前的人工智能，重点工作就是设计出强大的学习类算法。核心方法就是人工神经网络（当前更多地称深度神经网络）。



◆三、学习类算法及本质特征分析

◆典型的学习类算法及它们试图解决的问题。

◆1、非平稳（更严谨的表达应该是缓平稳！）信号的自适应滤波。

◆典型算法：LMS（Least-Mean-Square）算法和RLS（Recursive Least Square）算法。

◆目的是随着非平稳（缓平稳）信号的动态变化，计算最佳滤波器，一般是基于AR模型。然后对缓平稳数据中包含的信号进行“最佳”预测。

◆它们是线性均方估计器的推广。其中，学习的特点体现在：期望（导师）是预先给出的，通过动态地输入非平稳（缓平稳）信号，逐步训练滤波器系数。当逼近误差达到设定的最小值，或滤波器的系数变化很小，就停止训练，进入滤波应用阶段。

◆这应是最初级的有监督学习类算法。

◆这也应该是人工神经网络学习算法的一个重要起始点！



◆三、学习类算法及本质特征分析

◆典型的学习类算法及它们试图解决的问题。

◆2、回归分析：对数据进行建模

◆典型算法：线性多元回归分析/广义线性回归分析。

◆具有充分代表性的采样（统计特征完备的采样）是回归分析的基础。若数据没有“正确”的统计特征，回归分析结果谈不上什么正确性。（广义）线性假设也是基础，这可视为建模的先验假设，预先认为（广义）线性模型是合理的！

◆回归分析既确定（广义）线性模型本身，也确定最佳叠加系数。

◆训练数据与检验数据的采样充分性是基础保障；交叉验证试图保证所建模性的合理性，既不能过拟合，也不能欠拟合。

◆得到正确的回归模型后，就可利用这个模型进行预测等。

◆我认为：回归分析是理解（深入理解）学习类算法的、最好的切入算法。



◆三、学习类算法及本质特征分析

◆典型的学习类算法及它们试图解决的问题。

◆2、回归分析：对数据进行建模

- ◆这的确体现了计算机学习的基本方式，使得计算机具有了某种“智能”（尽管是很低级的智能！）。
- ◆人类的学习大致也是这种过程。不过，人类学习的方式更多样化，而且更多地利用概念进行学习。
- ◆Von Nuemann构型计算机目前只能从“数据”中进行学习，还不能从概念中学习。
- ◆今后的AI计算机性能还无法预测！
- ◆“专家系统”的学习方式，目前似乎被基于深度神经网络的学习方式击败了！



◆三、学习类算法及本质特征分析

◆典型的学习类算法及它们试图解决的问题。

◆3、分类问题

- ◆让计算机依据数据中蕴含的规律进行分类，是学习类算法的另一个典型。
- ◆训练出一个性能良好的分类器，首先还是要有统计特征完备的采样数据。
- ◆分类器算法设计是核心。SVM及各种变种、深度神经网络是公认的、典型的分类器。
- ◆**基于复杂数据的分类问题，绝对是一个强非线性问题。**强非线性问题的求解目前是没有完善解法的，只能在Bayes估计的思想下，尽可能设计出好算法。
- ◆把分类问题抽象为寻找最佳决策面的问题，又退化为对决策面的回归分析建模问题。当然，数据复杂时，这也是一个强非线性问题。



◆三、学习类算法及本质特征分析

◆典型的学习类算法及它们试图解决的问题。

◆3、分类问题

- ◆学习类算法的目的是做出合理的决策。这一定要具备一定的自主性，否则发展机器学习算法的必要性就值得怀疑了！
- ◆既然要让计算机根据数据进行学习，然后进行自主决策，必须思考这中间的根本逻辑。到底该如何做到此事，是否仅靠回归分析建模与解决分类问题就行了！
- ◆首先，还是与人的学习与决策过程进行类比，人的学习过程不是依赖于数据的，而是概念、模式和图像等。人脑决策机制也是异常复杂的，做出对一个专业问题的判断也很复杂，更不用讲对生活中问题的决策了。显然，计算机算法永远也不可能像人类进行决策！
- ◆我自己的体会：让计算机从数据中进行学习，学到的只能是数据中蕴含的结构（相关性），它们与决策目标之间存在某种线性（或非线性）的关系。找到这种关系后，计算机能进行一定程度的决策。



◆三、学习类算法及本质特征分析

◆典型的学习类算法及它们试图解决的问题。

◆3、分类问题

- ◆ 人类即便是进行专业问题的决策，我认为基本上也是基于模式的（概念也可转化为模式）。所谓抽象思维能力很强的人，无非是他/她的几何直观很强（指对物理、工程等专业问题，非历史、人文、哲学方面的问题）。而几何直观，就是模式！
- ◆ 尽管如此，让计算机找到数据中存在的模式，建立模式与决策目标之间的联系，然后进行决策，依然与人类的决策相差太远。计算机做决策考虑因素过于单一或过于线性化。人类决策绝对是非线性过程。我不认为计算机能做到与人脑相比拟的决策。
- ◆ 不过，计算机决策的特长在于它能充分地利用大数据（必须要是大数据）中蕴含的统计信息进行特定目标的判断，这是人类做不到的。
- ◆ 因此，计算机永远也不可能取代人类，但计算机“智能”有可能打败人类，打败人类并不需要完全拥有类人的智能。这一点正是很多忽悠大众的AI“传播”者不能讲明，或者不想讲明的要点。



◆三、学习类算法及本质特征分析

◆典型的学习类算法及它们试图解决的问题。

◆3、分类问题

◆基于上述认知，进行学习类算法研究时，就不会有困惑了。

◆尽力发展优秀的算法，让计算机从数据中挖掘出蕴含的统计信息，然后基于特定目的进行判决就够了！这与FWI并没什么区别，尽力把ML这件事做好就行！

◆学习类算法的最高目标当然是逼近类人的决策，这样，既发挥了计算机的特长，又兼顾了模拟人类的智能。至于该理想的目标能否实现只能走着看！神经网络是基本的探索路线。

◆千万不要把AI与人类智能混为一谈（目前二者是没什么关系的！）。

◆这样，我们在研究ML中各种算法时，就可根据自己的应用目的，随心所欲了！

◆既然对学习类算法的判断已经确定，如何开展对学习类算法的研究和应用？



◆三、学习类算法及本质特征分析

◆典型的学习类算法及它们试图解决的问题。

◆3、分类问题

◆既然对学习类算法的判断已经确定，如何开展对学习类算法的研究和应用？

目录

- ◆一、概述
- ◆二、Learning From data的数学基础
- ◆三、非学习类算法及本质特征分析
- ◆四、学习类算法及本质特征分析
- ◆五、关于学习类算法、ML及AI的思考
- ◆六、勘探地震数据分析中的ML算法探讨
- ◆七、总结与讨论



◆五、关于学习类算法、ML及AI的思考

◆首先，还是要理出学习类算法面对的本质问题是什么！

◆在信号/图像处理、参数估计中，我们已经给出如下核心问题：正问题和Bayes框架下的反问题。

◆学习类算法抽象后的核心问题是什么？有明确的定义吗？

◆我觉得有无定义首先要界定学习类算法想干什么或想达到什么目的！

◆若学习类算法的目的是：从数据中找到信息，根据找到的信息做决策，而不是更宏大的目的（即模仿人做决策）。那么，学习类算法的核心问题应是：

◆对数据进行建模，或抽取数据中包含的结构特征或模式，可把此对应于正问题。

◆建立模式与决策目标之间的关系，这是一个新问题，也应是学习类算法独有的！可把此对应为反问题。

◆解决决策过程中存在的各种问题，主要是决策失误、决策结果达不到要求的问题。



◆五、关于学习类算法、ML及AI的思考

◆据此给出学习类算法的定义：

◆学习类算法是从数据中提取信息，并能据此进行自主决策的一类算法。



◆五、关于学习类算法、ML及AI的思考

- ◆分类和聚类应是决策过程中的核心问题。
- ◆我相信：人类的决策基本也是如此！必须要回答譬如一件事可做、不可做；这是什么，这不是什么；等等诸如此类的问题。
- ◆人类的决策是计算机学不来的，但人类决策时总会给出一些原则性条款（或理由）。
- ◆学习类算法的决策，我认为首先也是建立决策原则，Bayes决策我依然认为是最合理的，用后验概率最大化来进行决策！
- ◆基于上述逻辑，可以设计出各种各样的学习类算法或机器学习算法。



◆五、关于学习类算法、ML及AI的思考

◆为什么当前深度神经网络会占主导呢？

- ◆仔细琢磨，这实际上是很让人困扰的！本来应在Bayes决策下，百花齐放的，为什么成了一枝独秀呢？
- ◆网格型算法本身并无什么新颖性，借用脑神经学科的发展，构造出神经网络算法，也没什么特别的，仿生优化算法是科学界的常识。
- ◆但神经网络算法，尤其是深度神经网络算法，在图像处理与语言处理中的成功，与CNN+U_Net表达图像与语音信号的模式的能力是大有关联的。
- ◆但必须从数学上讲明白其内蕴的数学逻辑，才能把网格型算法大大地向前推进。
- ◆对于非图像/语音类的数据，应该构建什么学习类的算法？（大规模的）参数估计问题适宜提成学习类算法吗？



◆五、关于学习类算法、ML及AI的思考

- ◆关于决策，实际上就是依据决策原则做出“最佳”决定。
 - ◆Bayes决策理论的确是一个很抽象、很美妙的理论框架，它可以提供思想指导，但它并不实用。
 - ◆模拟退火、遗传算法、增强学习、仿生类算法（蚁群追踪、粒子群算法等）。它们都在实现最佳决策，和Bayes决策理论的关系或明或暗。但我更偏爱Bayes决策理论，它给了我们一个抽象的、关于最佳决策是思维框架。
 - ◆各种分类算法更是在解决一定的决策问题。 Bayes分类器看起来更有普遍意义！

目录

- ◆ 一、概述
- ◆ 二、Learning From data的数学基础
- ◆ 三、非学习类算法及本质特征分析
- ◆ 四、学习类算法及本质特征分析
- ◆ 五、关于学习类算法、ML及AI的思考
- ◆ 六、勘探地震数据分析中的ML算法探讨
- ◆ 七、总结与讨论



◆六、勘探地震数据分析中的ML算法探讨

- ◆ 勘探地震中最核心的问题是成像问题、是参数估计问题。
 - ◆ 偏移成像问题是很难想象能提出学习类算法的，我个人认为也没什么必要性！
 - ◆ 大规模系统参数估计，我也不认为能由深度神经网络算法实现。
 - ◆ 不利用已有的波动方程，而是从数据中包含的统计规律进行大规模系统参数估计显然是愚蠢的选择。



◆六、勘探地震数据分析中的ML算法探讨

- ◆但是，由学习类算法提取出大规模系统参数估计时正则化所需的约束信息，尤其是结构信息是可以的，地质逻辑的合理性也可由学习类算法获得保障。
 - ◆去噪、插值这类问题，可以由学习类算法来完成，但必要性不是很大。
 - ◆拾取初至、速度谱解释这种需要决策的问题，发展学习类算法是必要的。
 - ◆地震剖面上层位、断层、特殊地质体的识别，发展学习类算法是必要的。
 - ◆凡是符合计算机自主决策的场景都可发展学习类算法。
- ◆对学习类算法，不可神化，不可迷信DL。
- ◆但是，只要是能由计算机依据数据进行自主决策的场景都可以发展出学习类算法。合适且必要的应用场景的选择很关键！

目录

- ◆一、概述
- ◆二、Learning From data的数学基础
- ◆三、非学习类算法及本质特征分析
- ◆四、学习类算法及本质特征分析
- ◆五、关于学习类算法、ML及AI的思考
- ◆六、勘探地震数据分析中的ML算法探讨
- ◆七、总结与讨论



◆六、总结与讨论

◆学习类算法的特点：

- ◆随着训练样本的增加，算法性能逐渐完善，泛化解决问题的能力逐渐增强。

◆学习类算法的目的：

- ◆通过从数据中的学习，拥有自主决策能力。

◆学习类算法的定义：

- ◆学习类算法是从数据中提取信息，并能据此进行自主决策的一类算法。



◆六、总结与讨论

◆发展学习类算法的要点：

- ◆对数据进行建模，或抽取数据中包含的结构特征或模式，然后建立模式与决策目标之间的关系。最后用于解决决策问题。

◆WPI的研究如何由非学习类的算法逐渐转移到二者并重，最后逐渐发展到以学习类算法为主？

- ◆进一步强化非学习类算法，重点是发展非线性、非高斯、非平稳假设下的数据分析算法；
- ◆改造非学习类算法，引入合理的决策机制，发展出新的学习类算法。
- ◆针对模拟人脑的人工神经网络，在网络构型和学习算法上提出自己的新理念和新做法。在地震数据分析和一般数据分析的应用中，获得好的应用效果。



谢谢
欢迎批评指正