

沟通技巧个人作业：文献综述

机器学习中多元线性回归算法的梯度下降求解及其应用

1.研究的背景与意义

机器学习是一门多领域的交叉学科，设计概率论，统计学，凸分析，算法复杂度理论等多种理论知识，专门用于研究计算机如何模拟和实现人类的学习行为用以获取新的知识和技能，将已有的知识结构进行重新的组织和整理，这项技术是人工智能技术的核心技术，是数据驱动的，使得计算机具有“智能”的根本途径。机器学习是人工智能及模式识别领域的共同研究热点，其理论和方法已被广泛应用于解决工程应用和科学领域的复杂问题。

机器学习中的两大任务分别是**回归和分类**，各种各样的回归算法用于完成机器学习中的一大任务---回归，简单来说就从一系列数据除法，确定某些变量之间的定量关系式，回归的目的是预测数值型的目标值，他的目标是接受连续的数据，寻找最适合数据的方程，并能够对特定值进行预测。这个方程称为回归方程，而求回归方程显然就是求该方程的回归系数，求这些回归系数的过程就是回归。

而线性回归则是诸多回归算法中形式最为简单的一类算法，最简单的情况下回归方程可以用 $y = ax + b$ 来表示，而一般在机器学习中用到的都是多元的线性回归模型，表示为 $h_{\theta}(x) = \sum \theta_i x_i = \theta^T x$ ，如果一个数据集有 n 个特征，那么 θ 就是一个 $n+1$ 维度的向量，而线性回归算法的根本目标也就是求解出合适的 θ ，通过定义损失函数 $f(\theta) = \frac{1}{m} \sum_{i=1}^n (\theta^T x^{(i)} - h_{\theta}(x^{(i)}))^2$ 来寻找使得损失函数最小的 $n+1$ 维向量 θ 的值，在这个过程中，梯度下降法是最常见的一种处理多元线性回归机器学习的方法。

本文综述将着重研究机器学习中的多元线性回归算法的梯度下降求解方法的发展情况和最新应用，研究这一问题有助于在机器学习算法发展遇到瓶颈，传统机器学习算法在深度学习算法火热的当下逐渐式微的情况下，为传统机器学习算法的发展寻找新的突破方向，同时也可以认识到传统机器学习方法自身存在的一些不足和值得改进之处。

2.研究的现状与存在的问题

2.1 梯度下降法的引入和发展

梯度下降法是使用最广泛的一种用于求解线性回归模型的方法，相比于最小二乘法求得的结果更加准确，梯度下降法通过计算损失函数并进行迭代的方式寻找拟合程度最高的向量 θ 的值，也就是使得损失函数值最小的点，一般分为批量梯度下降算法，随机梯度下降算法和小批量梯度的下降算法这三种实现框架。

确定学习率和选择优化的方向是梯度下降法的核心，近年来国内外的研究者取得了诸多成果，下面我们根据所阅读的文献中的内容来研究梯度下降法的最新研究进展。

2.1.1 目标函数优化

多元线性回归模型的损失函数往往是一个凸函数，这意味着这种函数只存在一个全局最小值，不存在局部的最小值，但是当目标函数非常复杂的时候，使用梯度下降法求解多元线性回归往往效果欠佳。

而在目标函数不是凸函数的情况下，**论文[1]**中作者Huo,Z和Huang.H提出了基于非凸优化方差约简的小批量梯度下降法，并对其进行了算法复杂度等方面的理论分析，异步的随机梯度下降法目前已经被广泛应用于机器学习的优化问题中，结果表明，当目标函数为凸函数的时候该技术具有线性的收敛速度。而对于目标函数不是凸函数的情况，作者采取了分布式内存架构和共享内存架构两种方式来实现异步的随机梯度下降算法来训练多元线性回归模型，并且证明了这两种方法的收敛速度都是 $O(\frac{1}{T^{0.5}})$ 。

此外，**论文[2]**中作者J.Flieg提出了一阶光滑的多目标函数优化方法，并且证明了这些方法具有一阶的全局收敛性，对于非凸函数，凸函数和强凸函数的全局速率和单目标优化的梯度下降率进行了理论分析，使得梯度下降法在训练多元线性回归模型时在多项式复杂度的时间内达到了零损失，这是因为训练过程中的矩阵运算中诱导出了Gram矩阵的特殊结构，这种特殊结构意味着梯度下降方法在全局上也是最优的。

2.1.2 基于动量的随机梯度下降算法优化

由于随机梯度下降算法的震荡性，目标的 θ 会在损失函数的最小值附近游走，在这种情况下，动量在随机梯度下降距离中加入上一次迭代的更新项，将其作为新的模型参数的下降距离，也就是：

$$\theta := \theta - (\gamma v_{t-1} + \alpha \delta_{\theta} MSE(\theta)) \quad (1)$$

这可以使得模型在更新时累积了之前积累的所有动量，使得梯度下降的速度越来越快。对于目标函数的非光滑优化问题，**论文[3]**中对于目标函数不光滑时候的多元线性回归问题，通过合理设置梯度下降的步长，证明了这种由Polyak提出的Heavy-ball动量方法具有最优的单个收敛速率，从而证明了这种动量优化方法将梯度下降的收敛速度提高了最优的情况。

而针对BP算法存在的问题，**文章[4]**在NAG动量优化的基础上提出了基于黄金比例的动量确定线性回归的加速梯度的策略，并应用于手写字体的识别，取得了比较好的测试结果，他的做法是：

$$\theta := \theta - (\gamma v_{t-1} + \alpha \delta_{\theta} MSE(\theta - \gamma v_{t-1})) \quad (2)$$

改进的随机方差消减梯度下降法(SVRG)可以解决SGD因受到数据噪声的干扰而收敛速度不理想的问题，**文章[5]**提出了一种基于SVRG算法的分布式算法topkSVRG使得该算法实现了线性的收敛速度。

我们可以将这些动量优化方法归结成如下的表格：

优化算法	算法的概述
动量优化	避免在局部最优上震荡
NAG	考虑上一次的梯度，在坡度上升之前减速
SVEG	在固定的学习率的情况下可以取得较快的收敛速度

2.2 基于梯度下降的多元线性回归算法的应用

2.2.1 基于多元线性回归方法的结果预测

多元线性回归方法最重要的作用之一依然是用于基于现有数据的预测分析，比如工业中的化工反应数据预测(如聚丙烯的熔融指数)，而**文章[6]**根据影响油田产量的重要因素，通过多元线性回归分析，建立了油田产量线性回归分析的预测模型，建立了油田年产量预测的多元线性回归模型，模型训练的过程中采用了梯度下降法，同时通过对回归参数求解过程中统计信息的分析，对模型进行优化，得到了产量预测的一个较优的结果。

2.2.2 基于多元线性回归的人脸识别

文章[7]提出了一种基于线性回归的人脸识别方法，最经典的基于多元线性回归的人脸识别算法是这样的：

Algorithm: Linear Regression Classification (LRC)

Inputs: Class models $\mathbf{X}_i \in \mathbb{R}^{q \times p_i}, i = 1, 2, \dots, N$ and a test image vector $\mathbf{y} \in \mathbb{R}^{q \times 1}$.

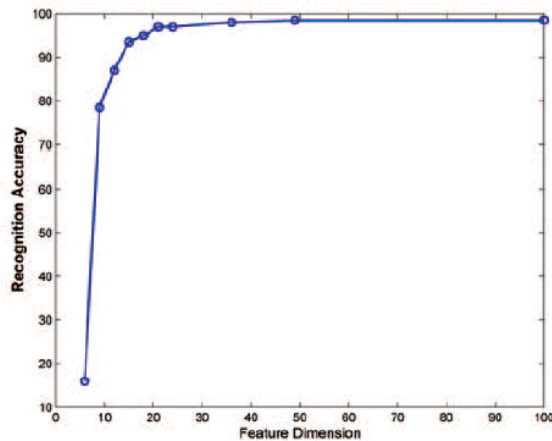
Output: Class of \mathbf{y}

1. $\hat{\beta}_i \in \mathbb{R}^{p_i \times 1}$ is evaluated against each class model,

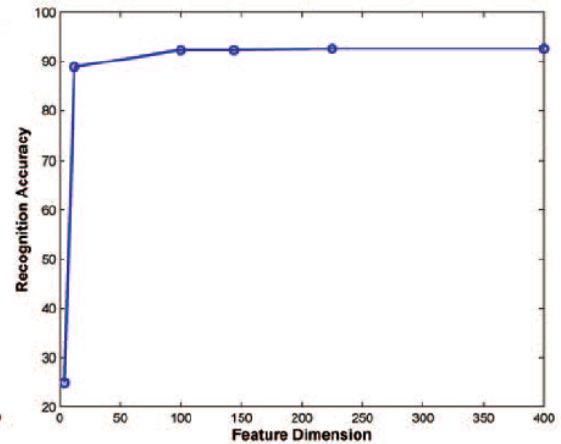
$$\hat{\beta}_i = (\mathbf{X}_i^T \mathbf{X}_i)^{-1} \mathbf{X}_i^T \mathbf{y}, \quad i = 1, 2, \dots, N$$
2. $\hat{\mathbf{y}}_i$ is computed for each $\hat{\beta}_i, \hat{\mathbf{y}}_i = \mathbf{X}_i \hat{\beta}_i, \quad i = 1, 2, \dots, N$
3. Distance calculation between original and predicted response variables $d_i(\mathbf{y}) = \|\mathbf{y} - \hat{\mathbf{y}}_i\|_2, \quad i = 1, 2, \dots, N$
4. Decision is made in favor of the class with the minimum distance $d_i(\mathbf{y})$

而本文中的该模型利用了单对象类的模式位于线性子空间的基本概念，用于探测图像表述为类特定库的线性组合，利用梯度下降方法求解逆问题，并用最小重构误差来判断识别的优劣程度，这种线性回归的分类算法属于最近邻子空间分类，该算法在多个标准数据上进行了广泛的评估，可以清楚地反应人脸识别的有效性，针对连续性遮挡的问题，提出了一种模块化的线性回归方法，引入了新的基于距离的证据融合(DEF)算法，在迄今为止所有的相关研究中取得了最好的结果。他们在AT&T实验室提供的AT&T人脸识别数据集上对他们所提出的算法进行了验证，结果如下：

Evaluation Protocol	Approach	Recognition Rate
EP1	Fisherfaces [24]	94.50%
	ICA [24]	85.00%
	Kernel Eigenfaces [24]	94.00%
	2DPCA [24]	96.00%
	ERE [27]	97.00%
	LRC	93.50%
EP2	Fisherfaces [24]	98.50%
	ICA [24]	93.80%
	Eigenfaces [24]	97.50%
	Kernel Eigenfaces [24]	98.00%
	2DPCA [24]	98.30%
	ERE_ \mathbf{S}^b [27]	99.25%
	ERE_ \mathbf{S}^t [27]	99.00%
	LRC	98.75%

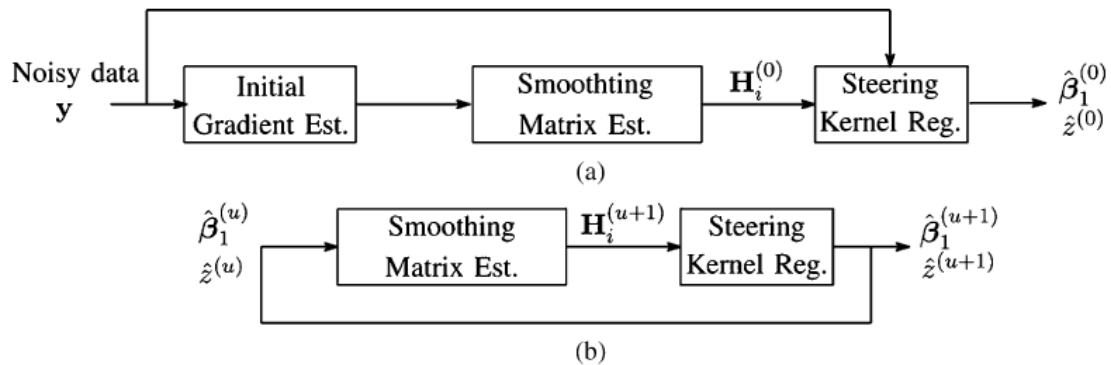


(a)



(b)

可以看到他们的算法训练的结果非常优秀，是多元线性回归模型在人脸识别中的一项重要应用成果。值得一提的是文章[8]还提出了一种核回归(Kernel Regression)的算法用于图像的去噪，缩放，插值和融合等操作，而核回归算法的核心思想依然是基于多元线性回归和梯度下降法，此外该文章还和一些流行的现有的方法建立了关键的联系，包括双边滤波器，非线性滤波器等，这种核回归算法的过程如下图所示：



3.总结与展望

本文主要概括性总结了梯度下降法在多元线性回归模型中的一些算法上的优化和实际应用，包括目标函数的优化和基于动量的随机梯度下降算法优化，实际应用场景包括多元线性回归算法的老本行结果预测和一种新的应用场景--人脸识别。我们可以看到多元线性回归作为最古老的机器学习算法之一在机器学习和人工智能领域的发展中依然历久弥新[9]。虽然目前传统机器学习领域的发展已经逐渐碰到了瓶颈，但是在阅读了上述文献之后，我依然感受到以多元线性回归为代表的经典机器学习算法将在学术研究领域和工业应用领域焕发新的光彩。

4.参考文献

- [1] Huo,Z.,Huang,H.Asynchronous mini-batch gradient descent with variance reduction for non-convex optimization[J]. USA:THIRTY-FIRST AAAI CONFERENCE ON ARTIFICIAL INTELLIGENCE, 2017:2043-2049.
- [2] J.Fliege,A.I.F.Vaz,L.N.Vicente(2019)Complexity of gradient descent for multiobjective optimization[J].Optimization Method and Software,2019,34(5):
- [3]:程禹嘉,陶蔚,刘宇翔,等.Heavy-Ball型动量方法的最优个体收敛速率[J].计算机研究与发展,2019,56(8):1686-1694.
- [4]: 景立森,丁志刚,郑树泉,等.基于NAG的BP算法的研究与改进[J].计算机应用与软件,2018,35(11):272-277.
- [6]: Liang Guo, Xianghui Deng, Application of Improved Multiple Linear Regression Method in Oilfield Output Forecasting, 2009 International Conference on Information Management, Innovation Management and Industrial Engineering
- [7]:Imran Naseem,Roberto Togneri, Senior Member, IEEE, and Mohammed Bennamoun,Linear Regression for Face Recognition, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 32, NO. 11,
- [8]: Hiroyuki Takeda, Student Member, IEEE, Sina Farsiu, Member, IEEE, and Peyman Milanfar, Senior Member, Kernel Regression for Image Processing and Reconstruction, IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 16, NO. 2, FEBRUARY 2007
- [9]:韩阳,孙佳泽,王昊天, 浅谈人工智能的发展历程及瓶颈[J].计算机教育, 2019第十期
- [10]:李兴怡,岳洋,梯度下降算法研究综述[J].软件工程,2020,2(2)

[11]:Kavitha S,Varuna S, Ramya R A Comparative Analysis on Linear Regression and Support Vector Regression, IEEE, IC-GET(2016)