摘要

差异隐私是最近的敏感数据计算框架，在大型数据集的制度中显示出相当大的前景。随机梯度方法是在数据丰富的方案中学习的流行方法，因为它们在计算上易于处理和可扩展。在本文中，我们推导出随机梯度下降的差异私有版本，并根据经验进行测试。我们的结果表明，标准SGD由于差异隐私而经历了高度可变性，但批量大小的适度增加可以显着提高性能。

一.导言

随着数据变得更容易以数字化格式获取和聚合，设计可以对这些数据进行操作的高效算法已经成为信号处理，机器学习和相关领域的核心挑战。这些数据通常可能是私人的或敏感的，例如医疗或财务记录。差异隐私[1]正迅速成为设计可以保证可量化的隐私级别算法的流行框架。 差分私有算法保证在单个个体数据中不同的两个数据库下算法输出的对数似然比小于α。对于小α，无论该个体是否参与数据集，对手对个人的推论都是相似的。

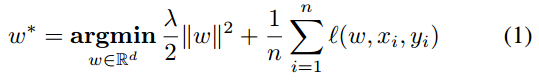
保证差别隐私涉及近似某些想要得到的算法或计算。近似性质会影响最终程序的性能或效用。例如，在参数估计中，保证隐私可以增加估计器的均方误差。解释这种影响的一种方式是目标效用级所需的样本大小随着隐私约束而增加。

在数据丰富的设置中，乍一看似乎学习算法既可以享受低隐私风险又可以享受高效用。但是，大数据集的优化方法也必须是可扩展的。随机梯度下降（SGD）算法最近受到了极大的关注，因为它们很简单并且满足与计算密集型学习方法相同的渐近保证[2]，[3]。但是，由于这些保证是渐近的，为了在有限数据集上获得合理的性能，从业者必须注意设置参数，例如更新的学习速率（步长）。为了在有限样本设置中减轻一些敏感性并提高SGD的性能，一些工作[4] - [6]建议将更新分组为“小批量”。 这可以在计算方面以适中的费用提高更新的稳健性，但是还将批量大小作为自由参数引入。

在本文中，我们推导出单点SGD和小批量SGD的差异私有版本，在真实和合成数据集上对它们进行评估。这些算法适用于一般凸目标的梯度下降 - 我们用逻辑回归分类说明了这种方法。我们证明差分私有单点SGD具有高方差，但批量大小适度增加可以显着提高性能。对于低维问题，私有算法的性能接近于非私有SGD。但是，我们表明批量大小可以提供多少帮助是有限的，并且性能取决于学习率。

二.准备工作

虽然我们提出的方法适用于一般优化方法，但我们将根据分类问题描述问题。在那里，数据是n个标记的例子，其中和。我们假设对于所有i，范数。在线性分类中，我们的目标是通过原点找到超平面，该超平面在很大程度上将标记为1的示例与标记为-1的示例分开。基于标记数据训练这种线性分类器的最流行的方法是通过求解正则化凸优化问题：



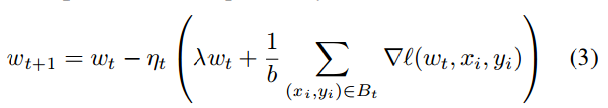
这里w是超平面分离器的法向量，是凸损失函数。 在机器学习文献中流行的选择是逻辑损失，这导致Logistic回归以及铰链损失，它导致支持向量机（SVM）。

SGD是用于求解1中的正则化凸优化问题的迭代算法.SGD以初始点开始，并且在步骤t，将迭代更新为：



.这里ηt是学习率，并且（sub）梯度是基于单个例子计算的。

在具有小批量更新的SGD中，代替单个示例，每个步骤t的更新基于大小为b的示例的小子集Bt。具体地，

。

这两种方法都是完全梯度更新的近似值 - 如果每个时间t的点从{1,2,…,n}均匀地采样，则每次迭代的期望梯度步长等于(1)中的全目标函数的梯度步长。更一般地，我们可以考虑具有凸损失函数的一般经验风险最小化。我们研究L2正则化目标，因为强凸性允许有利的理论保证。

我们的算法保证了差异隐私，这是一种由密码学驱动的隐私概念[1]。差异隐私在过去几年中在计算机科学界引起了极大的关注，并催生了越来越多的文献。隐私参数α> 0量化隐私风险;较低的α表示较高的隐私。

***定义1***：一个（随机化）算法A的输出位于域S中，如果对于所有S⊆S，则对于所有仅单个个体的值不同的数据集D和D'，则称为α差分私有，它是这种情况： 。

Dwork和Smith回顾了关于差别隐私的早期理论工作[7]。 Sarwate和Chaudhuri从信号处理的角度提供了一个教程[8]。 Le Ny和Pappas最近研究了差分隐私的信号处理方法[9]，[10]。与当前论文最相关的工作是差异私人分类[11]，[12]。杜奇等人提出了一种用于本地隐私的SGD方法[13]。随机梯度法是在线学习方法的一个例子。 Jain等人提出了另一种差异私人在线学习的方法[14];然而，他们的算法与批量正则化凸优化方法一样计算密集.PINQ [15]包使用噪声和操作来计算逻辑回归的完整噪声梯度步骤[16]。那里的目标是交换迭代的准确性。噪声感知器方法[17]也使用迭代噪声更新来学习分类器。我们在此关注步长和批量大小对SGD方法的影响，因此我们不会比较这些特定分类方法的性能。

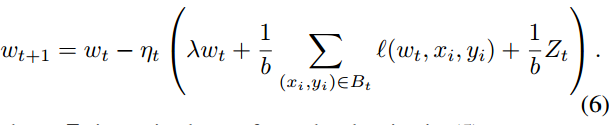
三.具有差别隐私的SGD

SGD更新的差异私有版本可写为：

，

其中每个Zt是独立的从密度：(5) 中绘制的中的随机噪声向量。

使用大小b的批量Bt的小批量更新的差异 - 私有版本可以写为：



，其中Zt也是从（5）中的密度中提取。

定理1表明，假设独立于敏感数据确定初始化点w0，批次Bt是不相交的，并且对于所有w,是有界的，这些更新是α差分私有的。

***定理1（SGD和小批量更新的隐私）***：假设我们用(6)中的小批量更新运行SGD,T个批次。如果选择初始化点w0独立于敏感数据，批次Bt是不相交的，并且如果对于所有w和,，则SGD与mini-批量更新是α差异私有的。

由于空间的限制，我们在这里提供了一个证明的草图，并推迟了这项工作的完整版本的细节。证明的关键思想是提供定理条件的观察，每次更新的全局敏感性是。现在通过将该观察结果与Dwork等人的结果相结合来证明\[1]，并且使用这样的事实，即由于批次中使用的样本是不相交的，因此隐私保证不会因批次而降低。

因为我们在每次迭代时添加噪声，所以SGD过程保证了“本地”意义上的差异隐私 - 每个人i可以选择αi，并且该方法可以通过调整Zt的分布来保证不同个体的不同级别αi的差异隐私。 Duchi等人也研究了略微不同的本地隐私概念[13]在统计设置中：算法可以从具有未知参数的分布中对个体进行采样，目标是估计参数。每次他们的算法都可以对新个体进行采样并接收嘈杂的次梯度估计。他们使用镜像下降来保证基于互信息标准的差异隐私变体下的隐私。

四.实验

A.数据集

我们考虑三个分类任务 - 一个在合成数据集上，两个在实际数据上。我们的合成数据集由从5维球体均匀绘制的n = 10000个样本组成，并且线性可分，边界为0.001。对于我们关于实际数据的第一个分类任务，我们使用KDDCup99数据集[18]，一个关于网络连接的入侵检测数据集。我们解决了正常与恶意分类任务，并使用大小为50000的子样本。对于我们关于实际数据的第二项任务，我们解决了MNIST数据集[19]上的“1对所有”分类任务，其中包括手写数字0到9的图像的60000个训练实例和1000个测试示例。在这两种情况下，我们通过规范化每个特征，将每一行投影到单位球，然后通过随机投影减少数据维度来预处理数据，这保留了差异隐私。对于KDDCup，我们使用d = 9的减小维度，对于MNIST，我们使用d = 15。

B.程序

我们使用SGD来训练逻辑回归模型。对于每次更新，我们使用（6）中的小批量更新批量大小，正则化参数λ= 0：0001和α= 1.在每种情况下，我们对整个训练数据进行单次传递。为了保持数值稳定性，在每次更新之后，我们将迭代wt投影到半径为1 =λ的球上。对于每个实验，我们研究了几种不同的设置学习率的方案。我们对训练数据的20多个随机排列以及私有算法噪声ZT的新随机样本的目标函数值进行了平均。误差线为单一标准偏差。由于我们对优化性能感兴趣，因此我们绘制了目标函数值 - 在未来的工作中，我们还将研究分类的准确性。

C.小批量减少差异

我们要问的第一个问题是SGD如何处理批量1，因为这是文献中研究最多的情况。图1的上半部分示出了MNIST数据集的目标值与用于学习速率ηt= 1 = pt的算法中的样本数量的关系。对于批量大小b = 1，差异私有SGD远离非私有目标并且还具有高方差。也就是说，每次迭代中添加的噪声都会阻止算法收敛。然而，如图的下半部分所示，适度的批量大小b = 10将差异私有SGD的方差减小到匹配非私有SGD的点，即使对于中等数量的数据点也是如此。

图1中的其他图表显示此行为也适用于KDDCup99数据集。尽管差分私有算法的方差缓慢减小，但选择b = 5使得小批量SGD性能几乎与非私有小批量SGD的性能相同。这两个实验表明，就客观价值而言，保证差别隐私可以使用具有中等批量的SGD来“免费”。我们在此强调所有这些示例都是低维问题，并且隐私参数α= 1.众所周知，差分私有学习算法通常具有与数据维度d线性对比并且与隐私成反比的样本复杂度。风险α。因此，α的适度减少或d的增加可能需要更多数据。有趣的是，在这些设置中，增加批量大小是否仍然可以使私有SGD与非私有SGD匹配。

D.选择适当的参数

我们的下一个实验是找到批量大小对这些算法性能的影响。图2显示了作为私有SGD，非私有SGD的批量大小的函数的目标值，以及使用所有数据点解决优化的集中式学习过程。在所有情况下，增加批量大小可以提高私有SGD的性能，但是存在一个限制 - 对于步长1 = pt，更大的批量大小实际上会降低性能。因为我们选择对数据进行单次传递以限制每次迭代的噪声，所以增加批量大小会减少迭代次数，因此对于每个问题都有最佳的b选择。学习率较大10 = pt时，较大批次的性能不会降低太多，并且目标的最终值更接近于集中式学习算法的最终值。

在强凸案例中对SGD的许多分析表明学习率ηt= 1 =λt保证了快速的收敛速度[2]。在我们的例子中，λ非常小，意味着物镜的凸度不是很强。为了了解为差异隐私添加的噪声的影响，我们模拟了b = 5的合成数据的两个学习率，1 = pt和1 =λt。图3中的结果表明选择快速减小的步长大大增加私人新元的差异。实际上，在随机近似方案中选择步长通常是艺术问题，差分私有噪声使这种选择复杂化。

五，结论

在本文中，我们研究了差别隐私如何影响小批量随机梯度下降（SGD）。当数据充足时，隐私是“负担得起的”，而SGD策略的计算效率更高。我们发现，在许多情况下，差异私有SGD的性能接近于非私有SGD的性能，尤其是批量较大的SGD。在随机优化中，通过将点组处理在一起，可以改善算法的可变性和保护隐私的噪声的影响。我们的实验表明，隐私影响最佳批量大小b和学习率ηt。我们的工作建议的一些有趣的未来方向包括：量化维度d和隐私参数α的影响，允许每个点的不同αi，并使用多次通过数据来权衡迭代，完全隐私损失（通过差异的合成结果）隐私）和错误。这些修改可以使差异私人学习在实际环境中更有效。

致谢

KC和SS感谢NIH U54-HL108460，Hellman Foundation和NSF IIS 1253942的支持。