随机优化问题

- 随机优化问题可以表示成以下形式：

$$\min_{x\in\mathcal{X}} \quad \mathbb{E}_\xi[F(x,\xi)] + h(x),$$

其中$\mathcal{X}\subseteq\mathbb{R}^n$表示决策变量$x$的可行域，$\xi$是一个随机变量．对于每个固定的$\xi$，$F(x,\xi)$表示样本$\xi$上的损失或者奖励．正则项$h(x)$用来保证解的某种性质．

- 变量ξ的数学期望$\mathbb{E}_\xi[F(x,\xi)]$一般是不可计算的．为了得到目标函数值的一个比较好的估计，在实际问题中往往利用ξ的经验分布来代替其真实分布．具体地，假设有N个样本$\xi_1,\xi_2,\cdots,\xi_N$，令$f_i(x)=F(x,\xi_i)$，得到优化问题

$$\min_{x\in\mathcal{X}} \quad f(x) \overset{\text{def}}{=\!=} \frac{1}{N}\sum_{i=1}^{N} f_i(x) + h(x), \tag{35}$$

并称其为经验风险极小化问题或者采样平均极小化问题．

- 该问题通常是难以求解的，一方面是因为样本数N比较多，另一方面是因为优化问题的可行域所在空间维数n比较大．

随机主成分分析

- 在主成分分析中，如果样本点ξ服从某个零均值分布$\mathcal{D}$，那么找方差最大的d维子空间的优化问题可以写成

$$\max_{X\in\mathbb{R}^{p\times d}} \quad trX^{\mathrm{T}}\mathbb{E}_{\xi\sim\mathcal{D}}[\xi\xi^{\mathrm{T}}]X \quad \text{s.t.} \quad X^{\mathrm{T}}X=I, \tag{36}$$

  其中$\mathbb{E}_{\xi\sim\mathcal{D}}[\xi\xi^{\mathrm{T}}]$为$\xi$的协方差矩阵.
- 在实际中，分布$\mathcal{D}$是未知的，已知的只是关于分布$\mathcal{D}$的采样．比如在在线主成分分析中，样本ξ_t是随着时间流逝依次获得的．这些已有的样本可以看作训练集.
- 随机主成分分析关心的问题是在求得问题(36)的高逼近解过程中需要的样本数量以及所消耗的时间．受制于计算机内存的限制，我们还需要考虑在有限内存情况下的逼近解的计算与分析.

分布式鲁棒优化

深度学习是机器学习的一个分支，通过利用神经网络来对数据进行表征学习．深度学习的目的是从已有的未知分布的数据中学出一个好的预测器，其对应优化问题

$$\min_{h} \quad \mathbb{E}_z[F(h,z)],$$

其中预测器h是优化变量，并对应于神经网络的参数．

- 因为数据z的真实分布的未知性，我们只有有限的样本点$z_1, z_2, \cdots, z_n$．在实际中的一种做法是将这些样本点对应的离散经验分布作为z的真实分布，对应的目标函数写成相应的有限和的形式，这种方式往往保证了在已有样本点上的高预测准确率．
- 但当我们拿到一个新的样本点时，该预测器的准确率可能会下降很多，甚至给出不合理的预测结果．即预测器的泛化能力较差．

分布式鲁棒优化

- 为了提高预测器的泛化能力，另外一种常用的方法是考虑分布式鲁棒优化问题

$$\min_{h} \quad \max_{\hat{z}\in\Gamma} \mathbb{E}_{\hat{z}}[F(h,\hat{z})],$$

这里集合Γ 中的随机变量的分布与真实数据的分布在一定意义下非常接近.

- 具体地，在选取Γ 时，我们需要考虑其对应的实际意义、可解性和数值表现. 给定数据的经验分布，一种方式是通过分布之间的熵来定义分布间的距离，从而定义Γ 为与经验分布的距离小于给定数的分布的集合. 目前常用的另外一种方式是利用分布间的Wasserstein 距离，这种距离的好处是其可以改变原来经验分布的支撑集.