机器学习 课程笔记

酥雨

zusuyu@stu.pku.edu.cn

June 17, 2022

目录

1 Inequalities & Concentration Bounds

定理 1.1 (Markov Inequality). 如果非负随机变量 X 期望存在, 则对于任意 $k>0$,

$$\mathbb{P}[X \geqslant k] \leqslant \frac{\mathbb{E}[X]}{k}$$

进一步地, 如果 r 阶矩 $\mathbb{E}[X^r]$ 存在, 则对于任意 $k>0$,

$$\mathbb{P}[X \geqslant k] \leqslant \min_{j \leqslant r} \frac{\mathbb{E}\left[X^j\right]}{k^j}$$

定理 1.2 (Chebyshev Inequality). 如果随机变量 X 方差存在, 则对于任意 $\varepsilon>0$,

$$\mathbb{P}[|X-\mathbb{E}[X]| \geqslant \varepsilon] \leqslant \frac{\operatorname{Var}[X]}{\varepsilon^2}$$

定义 1.3 (矩生成函数, Moment Generating Function, MGF). 如果随机变量 X 的任意 $n \in \mathbb{N}$ 阶矩存在, 则定义其矩生成函数为

$$M_X(t)=\mathbb{E}\left[\mathrm{e}^{tX}\right]=\sum_{i \geqslant 0} t^i \frac{\mathbb{E}\left[X^i\right]}{i!}$$

定理 1.4 (Chernoff Inequality).

$$\mathbb{P}[X \geqslant k] \leqslant \inf_{t>0} \mathrm{e}^{-tk} M_X(t)$$

接下来我们提出三个逐渐增强的定理, 从而最终证明 Chernoff Bound.

定理 1.5. $X_1, X_2, \cdots, X_n \sim$ i.i.d. $\mathcal{B}(1, p)$, 对于任意 $\varepsilon>0$,

$$\mathbb{P}\left[\frac{1}{n} \sum_{i=1}^n X_i-p \geqslant \varepsilon\right] \leqslant \mathrm{e}^{-n D_B(p+\varepsilon \| p)}$$

其中 $D_B(p \| q)$ 是两个 Bernoulli distribution $P=(p, 1-p), Q=(q, 1-q)$ 之间的相对熵.

证明.

$$\begin{aligned}
\mathbb{P}\left[\frac{1}{n} \sum_{i=1}^n X_i-p \geqslant \varepsilon\right] &= \mathbb{P}\left[\sum_{i=1}^n X_i \geqslant n(p+\varepsilon)\right] \\
&\leqslant \inf_{t>0} \mathrm{e}^{-tn(p+\varepsilon)} \mathbb{E}\left[\mathrm{e}^{t \sum_{i=1}^n X_i}\right] \\
&= \inf_{t>0} \mathrm{e}^{-tn(p+\varepsilon)} \prod_{i=1}^n \mathbb{E}\left[\mathrm{e}^{tX_i}\right] \\
&= \inf_{t>0} \mathrm{e}^{-tn(p+\varepsilon)}\left(p \mathrm{e}^t+1-p\right)^n \\
&= \inf_{t>0}\left(\frac{p \mathrm{e}^t+1-p}{\mathrm{e}^{t(p+\varepsilon)}}\right)^n
\end{aligned}$$

通过“简单”求导, 取 $t=\ln \frac{(1-p)(p+\varepsilon)}{p(1-p-\varepsilon)}$ 时上式右边取最小值, 从而有

$$\begin{aligned}
\mathbb{P}\left[\frac{1}{n} \sum_{i=1}^n X_i-p \geqslant \varepsilon\right] &\leqslant \left(\frac{\frac{(1-p)(p+\varepsilon)}{1-p-\varepsilon}+1-p}{\left(\frac{(1-p)(p+\varepsilon)}{p(1-p-\varepsilon)}\right)^{p+\varepsilon}}\right)^n=\left(\frac{\frac{1-p}{1-p-\varepsilon}}{\left(\frac{(1-p)(p+\varepsilon)}{p(1-p-\varepsilon)}\right)^{p+\varepsilon}}\right)^n \\
&= \left(\left(\frac{p}{p+\varepsilon}\right)^{p+\varepsilon}\left(\frac{1-p}{1-p-\varepsilon}\right)^{1-p-\varepsilon}\right)^n=\mathrm{e}^{-n D_B(p+\varepsilon \| p)}
\end{aligned}$$

□

定理 1.6. $X_1, X_2, \cdots, X_n \in [0,1]$ 是 n 个期望相同的独立随机变量, $\mathbb{E}[X_i] = p$, 对于任意 $\varepsilon > 0$,

$$\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n} X_i - p \geqslant \varepsilon\right] \leqslant \mathrm{e}^{-nD_B(p+\varepsilon\|p)}$$

证明. 注意到指数函数是下凸的, 根据 Jensen Inequality, 有

$$\mathbb{E}\left[\mathrm{e}^{tX}\right] \leqslant \mathbb{E}\left[X\mathrm{e}^t + (1-X)\mathrm{e}^0\right] = p\mathrm{e}^t + 1 - p$$

从而

$$\mathbb{E}\left[\mathrm{e}^{t\sum_{i=1}^{n} X_i}\right] \leqslant (p\mathrm{e}^t + 1 - p)$$

沿用定理 1.5 的证明即可. □

定理 1.7 (Chernoff Bound). $X_1, X_2, \cdots, X_n \in [0,1]$ 是 n 个独立随机变量, $\mathbb{E}[X_i] = p_i$, 记 $p = \frac{1}{n}\sum_{i=1}^{n} p_i$, 对于任意 $\varepsilon > 0$,

$$\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n} X_i - p \geqslant \varepsilon\right] \leqslant \mathrm{e}^{-nD_B(p+\varepsilon\|p)}$$

证明. 注意到对数函数是上凸的, 从而函数 $f(x) = \ln(x\mathrm{e}^t + 1 - x)$ 也是上凸的, 同样根据 Jensen Inequality, 有

$$\frac{1}{n}\sum_{i=1}^{n} \ln(p_i\mathrm{e}^t + 1 - p_i) \leqslant \ln(p\mathrm{e}^t + 1 - p)$$

从而

$$\mathbb{E}\left[\mathrm{e}^{t\sum_{i=1}^{n} X_i}\right] \leqslant \prod_{i=1}^{n}(p_i\mathrm{e}^t + 1 - p_i) \leqslant (p\mathrm{e}^t + 1 - p)^n$$

□

定理 1.8 (Additive Chernoff Bound). $X_1, X_2, \cdots, X_n \in [0,1]$ 是 n 个独立随机变量, $\mathbb{E}[X_i] = p_i$, 记 $p = \frac{1}{n}\sum_{i=1}^{n} p_i$, 对于任意 $\varepsilon > 0$,

$$\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n} X_i - p \geqslant \varepsilon\right] \leqslant \mathrm{e}^{-2n\varepsilon^2}$$

证明. 只需要证明 $D_B(p+\varepsilon\|p) \geqslant 2\varepsilon^2$ 即可. 听说可以暴力求导. □

定理 1.9 (Hoeffding Bound). $X_1, X_2, \cdots, X_n$ 是 n 个独立随机变量, $X_i \in [a_i, b_i]$, 记 $p = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{1}{n}\sum_{i=1}^{n}\frac{a_i+b_i}{2}$, 对于任意 $\varepsilon > 0$,

$$\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n} X_i - p \geqslant \varepsilon\right] \leqslant \mathrm{e}^{-\frac{2n\varepsilon^2}{\left(\frac{1}{n}\sum_{i=1}^{n} b_i - a_i\right)^2}} \leqslant \mathrm{e}^{-\frac{2n^2\varepsilon^2}{\sum_{i=1}^{n}(b_i-a_i)^2}}$$

定理 1.10 (McDiarmid Inequality). $X_1, X_2, \cdots, X_n \in \mathcal{X}$ 是 n 个独立随机变量, 如果对于 $f : \mathcal{X}^n \to \mathbb{R}$ 存在常数 $c_1, c_2, \cdots, c_n$ 使得

$$|f(x_1, \cdots, x_i, \cdots, x_n) - f(x_1, \cdots, x_i', \cdots, x_n)| \leqslant c_i$$

对于任意 $i \in [n], x_1, \cdots, x_n, x_i'$ 成立, 则对于任意 $\varepsilon > 0$, 有

$$\mathbb{P}\left[f(x_1, \cdots, x_n) - \mathbb{E}\left[f(x_1, \cdots, x_n)\right] \geqslant \varepsilon\right] \leqslant \exp\left(\frac{-2\varepsilon^2}{\sum_{i=1}^{n} c_i^2}\right)$$

定理 1.11 (Draw with/without Replacement). 有 m 个数 $a_1, \cdots, a_m \in \{0,1\}$, 记 $p = \frac{1}{m}\sum_{i=1}^{m} a_i$. $X_1, \cdots, X_n$ 为从 $\{a_1, \cdots, a_m\}$ 中的随机放回抽样, $Y_1, \cdots, Y_n$ 为从 $\{a_1, \cdots, a_m\}$ 中的随机不放回抽样, 则对于任意 $\varepsilon > 0$ 有

$$\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n} X_i - p \geqslant \varepsilon\right] \leqslant \mathrm{e}^{-2n\varepsilon^2}, \qquad \mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n} Y_i - p \geqslant \varepsilon\right] \leqslant \mathrm{e}^{-2n\varepsilon^2}$$

证明. 对于随机放回抽样, 显然每次抽样是独立的, 从而结论是 Chernoff Bound 的平凡推论.

对于随机不放回抽样, 注意到 $\mathbb{E}\left[\prod_{i\in I} Y_i\right] \leqslant \mathbb{E}\left[\prod_{i\in I} X_i\right]$ 对任意指标集 $I \subseteq \{1, \cdots, n\}$ 成立, 从而可以证明 $\mathbb{E}\left[\mathrm{e}^{t\sum_{i=1}^{n} Y_i}\right] \leqslant \mathbb{E}\left[\mathrm{e}^{t\sum_{i=1}^{n} X_i}\right]$. □

2 VC Theory

对一个分类器 f, 通常有两种评价指标: training error $err_S(f)=\mathbb{P}_{(x,y)\in S}[y\neq f(x)]$ 与 generalization error $err_D(f)=\mathbb{P}_{(x,y)\sim D}[y\neq f(x)]$. 接下来可能会不加声明地用 S 表示从数据集 D 中 sample 出来的训练集.

称 $err_D(f)-err_S(f)$ 为分类器 f 的 generalization gap. 我们提出**一致收敛 (uniformly converge)** 的概念, 它表示随着训练集 S 的增大, hypothesis space $\mathcal{F}$ 中的所有分类器 f 的 generalization gap 都会"一致"地被 bound 住.

定理 2.1 (Uniform Convergence when $|\mathcal{F}|<\infty$). S 是从数据集 D 中随机采样的训练集, $|S|=n$, 有

$$\mathbb{P}\left[\forall f\in\mathcal{F}, err_D(f)-err_S(f)\geqslant\varepsilon\right]\leqslant|\mathcal{F}|\mathrm{e}^{-2n\varepsilon^2}$$

证明. 对于某个确定的 $f\in\mathcal{F}$, 注意到 $err_S(f)=\frac{1}{n}\sum_{i=1}^n[y_i\neq f(x_i)]$, $\mathbb{E}\left[y_i\neq f(x_i)\right]=err_D(f)$, 故根据 Chernoff Bound 有 $\mathbb{P}\left[err_D(f)-err_S(f)\geqslant\varepsilon\right]\leqslant\mathrm{e}^{-2n\varepsilon^2}$. 再结合 Union Bound 即得结论. □

定理 2.2 (VC Theorem). 对于 VC-dimension (会在接下来定义) 为 d 的 hypothesis space $\mathcal{F}$, 从数据集 D 中随机采样大小为 n 的训练集 S, 则

$$\mathbb{P}\left[\sup_{f\in\mathcal{F}}|err_D(f)-err_S(f)|\geqslant\varepsilon\right]\leqslant4\left(\frac{2\mathrm{e}n}{d}\right)^d\mathrm{e}^{-n\varepsilon^2/8}$$

或者等价地, 有至少 $1-\delta$ 的概率, 对所有 $f\in\mathcal{F}$ 有

$$err_D(f)\leqslant err_S(f)+O\left(\sqrt{\frac{d\ln n+\ln(1/\delta)}{n}}\right)$$

为了接下来的叙述方便, 我们引入一些记号:

- 对于分类器 $f\in\mathcal{F}$ 以及数据点 $z=(x,y)\sim D$, 定义 $\phi_f(z)=\mathbb{1}[y\neq f(x)]$, 即每个 ϕ_f 是一个"长度为 $|D|$"的 01 串, 1 表示 f 会在这一位对应的数据点上出错.
- 定义 $\Phi_{\mathcal{F}}=\{\phi_f|f\in\mathcal{F}\}$. 由于以下不会超过一个 hypothesis space, 故省略角标简记为 Φ.

如此一来, 对于 $S=\{z_1=(x_1,y_1),\cdots,z_n=(x_n,y_n)\}$, 两种错误率 $err_S(f)$ 和 $err_D(f)$ 就分别等价于 $\frac{1}{n}\sum\limits_{i=1}^n\phi_f(z_i)$ 和 $\mathbb{E}_{z\sim D}\phi_f(z)$, 而我们需要限制的概率也变成了

$$\mathbb{P}_{S\sim D^n}\left[\sup_{\phi\in\Phi}\left|\frac{1}{n}\sum_{i=1}^n\phi(z_i)-\mathbb{E}_{z\sim D}[\phi(z)]\right|\geqslant\varepsilon\right]$$

引理 2.3 (Double Sampling). 取 $n\geqslant\frac{\ln2}{\varepsilon^2}$, 有

$$\mathbb{P}_{S\sim D^n}\left[\sup_{\phi\in\Phi}\left|\frac{1}{n}\sum_{i=1}^n\phi(z_i)-\mathbb{E}_{z\sim D}[\phi(z)]\right|\geqslant\varepsilon\right]\leqslant2\mathbb{P}_{S\sim D^{2n}}\left[\sup_{\phi\in\Phi}\left|\frac{1}{n}\sum_{i=1}^n\phi(z_i)-\frac{1}{n}\sum_{i=n+1}^{2n}\phi(z_i)\right|\geqslant\frac{\varepsilon}{2}\right]$$

通过 Double Sampling, 我们只需要限制 $\frac{1}{n}\sum\limits_{i=1}^n\phi(z_i)$ 与 $\frac{1}{n}\sum\limits_{i=n+1}^{2n}\phi(z_i)$ 的差. 考虑一种新的抽样方式, 先随机抽取 $\{z_1,\cdots,z_{2n}\}$, 再对其随机排列, 这样显然是与原先等价的, 即

$$\mathbb{P}_{S\sim D^{2n}}\left[\sup_{\phi\in\Phi}\left|\frac{1}{n}\sum_{i=1}^n\phi(z_i)-\frac{1}{n}\sum_{i=n+1}^{2n}\phi(z_i)\right|\geqslant\varepsilon\right]=\mathbb{E}_{S\sim D^n}\left[\mathbb{P}_\sigma\left[\sup_{\phi\in\Phi}\left|\frac{1}{n}\sum_{i=1}^n\phi(z_{\sigma(i)})-\frac{1}{n}\sum_{i=n+1}^{2n}\phi(z_{\sigma(i)})\right|\geqslant\varepsilon\right]\right]$$

这么做的意义是什么? 意义是可以先只考虑内层的 $\mathbb{P}_\sigma$ 而不管 $S\sim D^n$ 的选取. 看似强行取的随机排列 σ 是为了内层可以被 bound, 不然 $\mathbb{1}\left[\left|\frac{1}{n}\sum_{i=1}^n\phi(z_i)-\frac{1}{n}\sum_{i=n+1}^{2n}\phi(z_i)\right|\geqslant\varepsilon\right]$ 还不太方便处理.

记 $N^\Phi(z_1,\cdots,z_n)$ 表示 $\#\{(\phi(z_1),\cdots,\phi(z_n))|\phi\in\Phi\}$, 即 Φ 中的所有 01 串在数据点 $z_1,\cdots,z_n$ 上有多少种不同的. 从这个角度想, 其实 $\sup_{\phi\in\Phi}$ 只是在对有限项求 max, 故根据 Union Bound 可以得到

$$\mathbb{P}_\sigma\left[\sup_{\phi\in\Phi}\left|\frac{1}{n}\sum_{i=1}^{n}\phi(z_{\sigma(i)})-\frac{1}{n}\sum_{i=n+1}^{2n}\phi(z_{\sigma(i)})\right|\geqslant\varepsilon\right]\leqslant N^\Phi(z_1,\cdots,z_{2n})\mathbb{P}_\sigma\left[\left|\frac{1}{n}\sum_{i=1}^{n}\phi(z_{\sigma(i)})-\frac{1}{n}\sum_{i=n+1}^{2n}\phi(z_{\sigma(i)})\right|\geqslant\varepsilon\right]$$

其实这里写得不太严谨, 右边应该是对 $N^\Phi(z_1,\cdots,z_{2n})$ 个不同的 ϕ 分别求概率再相加, 但我们接下来会对任意 ϕ 限制 $\mathbb{P}_\sigma\left[\left|\frac{1}{n}\sum\limits_{i=1}^{n}\phi(z_{\sigma(i)})-\frac{1}{n}\sum\limits_{i=n+1}^{2n}\phi(z_{\sigma(i)})\right|\geqslant\varepsilon\right]$, 所以应该也无伤大雅.

对于一个特定的 $\phi\in\Phi$, 考虑 $\frac{1}{n}\sum\limits_{i=1}^{n}\phi(z_{\sigma(i)})$ 其实就是在 $\{\phi(z_1),\cdots,\phi(z_{2n})\}$ 这 $2n$ 个数中做不放回抽样, 故根据定理 1.11, 有

$$\begin{aligned}
\mathbb{P}_\sigma\left[\left|\frac{1}{n}\sum_{i=1}^{n}\phi(z_{\sigma(i)})-\frac{1}{n}\sum_{i=n+1}^{2n}\phi(z_{\sigma(i)})\right|\geqslant\varepsilon\right]&=2\mathbb{P}_\sigma\left[\frac{1}{n}\sum_{i=1}^{n}\phi(z_{\sigma(i)})-\frac{1}{n}\sum_{i=n+1}^{2n}\phi(z_{\sigma(i)})\geqslant\varepsilon\right]\\
&=2\mathbb{P}_\sigma\left[\frac{1}{n}\sum_{i=1}^{n}\phi(z_{\sigma(i)})-\frac{1}{2n}\sum_{i=1}^{2n}\phi(z_{\sigma(i)})\geqslant\frac{\varepsilon}{2}\right]\\
&=2\mathbb{P}_\sigma\left[\frac{1}{n}\sum_{i=1}^{n}\phi(z_{\sigma(i)})-p\geqslant\frac{\varepsilon}{2}\right]\\
&\leqslant 2\mathrm{e}^{-\frac{n\varepsilon^2}{2}}
\end{aligned}$$

从而我们得到了

$$\mathbb{P}_{S\sim D^{2n}}\left[\sup_{\phi\in\Phi}\left|\frac{1}{n}\sum_{i=1}^{n}\phi(z_i)-\frac{1}{n}\sum_{i=n+1}^{2n}\phi(z_i)\right|\geqslant\varepsilon\right]\leqslant 2\mathrm{e}^{-\frac{n\varepsilon^2}{2}}\mathbb{E}_{S\sim D^{2n}}\left[N^\Phi(z_1,\cdots,z_{2n})\right]$$

于是我们只需要限制 **Growth Function** $N^\Phi(n)=\max_{S\sim D^n}N^\Phi(z_1,\cdots,z_n)$ 即可. 除去 $N^\Phi(n)\equiv 2^n$ 这种平凡的情况 (这种情况意味着这种场景是 somehow not learnable 的), 我们指出 $N^\Phi(n)$ 是多项式增长的.

引理 2.4. 假设 $N^\Phi(d+1)<2^{d+1}$ 成立, 则对于任意 $n\geqslant d+1$, 都有 $N^\Phi(n)\leqslant\sum\limits_{k=0}^{d}\binom{n}{k}$.

证明. $N^\Phi(d+1)<2^{d+1}$ 说明存在一种 $w\in\{0,1\}^{d+1}$ 无法被 Φ 表示, 我们将其称为 **forbidden pattern**. 对于 $n\geqslant d+1$, 考虑指标集 $\mathcal{I}=\{i_1,i_2,\cdots,i_{d+1}\}\subseteq[n]$, 用 $w_\mathcal{I}\in\{0,1,*\}^n$ 表示在 $\mathcal{I}$ 指标上填 w, 其余位置填通配符 $*$ 得到的 n 位 01 串模式. 用 $E(w_\mathcal{I})$ 表示能被 $w_\mathcal{I}$ 模式匹配的 n 位 01 串集合, 我们需要做的是限制 $|\bigcup_\mathcal{I}E(w_\mathcal{I})|$ 的上界 (从而限制其补集的下界).

如果 $w=0^{d+1}$, 那么这个问题是好办的, 因为相当于 $\bigcup_\mathcal{I}E(w_\mathcal{I})$ 中包含了所有有至少 $d+1$ 个 0 的 01 串, 答案就恰好是 $\sum\limits_{k=d+1}^{n}\binom{n}{k}$. 如果 $w\neq 0^{d+1}$, 直观来看 $\bigcup_\mathcal{I}E(w_\mathcal{I})$ 的大小不会减小, 因此结论依然成立. 详细证明则是对于每个 $i\in[n]$, 把所有 $w_\mathcal{I}$ 在这一位上的 1 变成 0, 验证并不会使集合大小变大. $\square$

推论 2.5. 考虑 $\sum\limits_{k=0}^{d}\binom{n}{k}\leqslant\left(\frac{\mathrm{e}n}{d}\right)^d=O(n^d)$, 我们最终得到了

$$\mathbb{P}_{S\sim D^n}\left[\sup_{\phi\in\Phi}\left|\frac{1}{n}\sum_{i=1}^{n}\phi(z_i)-\mathbb{E}_{z\sim D}[\phi(z)]\right|\geqslant\varepsilon\right]\leqslant 4\left(\frac{2\mathrm{e}n}{d}\right)^d\mathrm{e}^{-\frac{n\varepsilon^2}{8}}$$

其中 $d=\max\{n|N^\Phi(n)=2^n\}$ 被定义为 hypothesis space $\mathcal{F}$ 的 VC dimension.

3 Game Theory

Game theory is the study of mathematical models of strategic interactions among rational agents, cited from Wikipedia.

我们引入“双人矩阵博弈”作为对博弈论最基础的介绍. 注意, 接下来我们考虑的所有问题都是零和的.

定义 3.1 (**Two-player Matrix Game**). 有一个 $M \in \mathbb{R}^{m\times n}$ 的矩阵. 两名玩家 Alice 和 Bob 参加了这场博弈. Alice, **the row player** 选择一行 $i \in [m]$, 相应的, Bob, **the column player** 选择一列 $j \in [n]$, 此时 Alice 获得收益 $-M_{ij}$, Bob 获得收益 M_{ij}.

我们首先探讨**纯策略 (pure strategy)** 的情景, 指的是 Alice 和 Bob 必须分别选择某个确定的行或列.

当 Alice 先做出选择时, 当她选出第 i 行后, 她会认为 Bob 会选择第 $j_i = \arg\max_j M_{ij}$ 列, 因此她会选择第 $\arg\min_i \max_j M_{ij}$ 行, 导致最终的博弈结果为 $\min_i \max_j M_{ij}$.

同理, 当 Bob 先做选择时, 他会选择第 $\arg\max_j \min_i M_{ij}$ 列, 导致最终的博弈结果为 $\max_j \min_i M_{ij}$.

我们指出在纯策略的情境下, 后手是有优势的, 即

定理 3.2. $\min_i \max_j M_{ij} \geqslant \max_j \min_i M_{ij}$ 对于任意 $M \in \mathbb{R}^{m\times n}$ 都成立, 同时存在 M', 使 $\min_i \max_j {M'}_{ij} > \max_j \min_i {M'}_{ij}$.

证明. 记 $i_0 = \arg\min_i \max_j M_{ij}, j_0 = \arg\max_j \min_i M_{ij}$, 有

$$\min_i \max_j M_{ij} = \max_j M_{i_0 j} \geqslant M_{i_0 j_0} \geqslant \min_i M_{i j_0} = \max_j \min_i M_{ij}$$

考虑 $M' = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$, 有 $\min_i \max_j {M'}_{ij} = -1, \max_j \min_i {M'}_{ij} = 1$. □

接着我们研究**混合策略 (mixed strategy)**, 其意味着玩家做出的决策可以不是确定的行列选择, 而是一个概率分布. 相应地, 得到的收益也就变成了期望收益.

形式化地, Alice 选择给出概率分布 $p = (p_1, \cdots, p_m) \in [0,1]^m$, Bob 给出概率分布 $q = (q_1, \cdots, q_n) \in [0,1]^n$. 合法的概率分布需要满足 $\|p\|_1 = \|q\| = 1$, 而此时两人的收益也分别是 $-p^{\mathrm{T}}Mq$ 与 $p^{\mathrm{T}}Mq$.

与纯策略的情境同理, 当 Alice 先手时, 博弈结果为 $\min_{p\in[0,1]^m, \|p\|_1=1} \max_{q\in[0,1]^n, \|q\|_1=1} p^{\mathrm{T}}Mq$, 当 Bob 先手时, 博弈结果为 $\max_{q\in[0,1]^n, \|q\|_1=1} \min_{p\in[0,1]^m, \|p\|_1=1} p^{\mathrm{T}}Mq$. 在接下来的叙述中, 我们默认 p, q 应取合法的概率分布, 而忽略在 $\min, \max$ 记号下的明确限制.

我们想要知道混合策略下后手还有没有优势. John von Neuman 告诉我们, 没有.

定理 3.3 (**von Neuman Minimax Theorem**).

$$\min_p \max_q p^{\mathrm{T}}Mq = \max_q \min_p p^{\mathrm{T}}Mq$$

4 Lagrange Duality, Linear Seperation, SVM and more

开摆了.

5 Boosting

Boosting 做的事情就是把一堆 classifier 放在一起, 从而得到一个更准确的 classifier.

以下算法中, γ-weak learning algorithm 表示其可以对于任意输入的带权训练集 D, 都能给出一个 classifier, 能在至少 $\frac{1+\gamma}{2}$ 的数据上得到正确的结果.

Algorithm 1 AdaBoost

Require: training set $S = \{(x_1, y_1), \cdots, (x_n, y_n)\}$, γ-weak learning algorithm $\mathcal{A}$

$D_1(i) \leftarrow \frac{1}{n}, \forall i \in [n]$.

for $t = 1 \to T$ **do**

  Use $\mathcal{A}$ to learn a classifier h_t based on D_t.

  $\varepsilon_t \leftarrow \sum\limits_{i=1}^{n} D_t(i)\mathbb{1}[y_i \neq h_t(x_i)]$

  $\gamma_t \leftarrow 1 - 2\varepsilon_t$ ▷ $\gamma_t \geqslant \gamma$

  $\alpha_t \leftarrow \frac{1}{2}\ln\frac{1+\gamma_t}{1-\gamma_t}$

  $Z_t \leftarrow \sum_i D_t(i)\exp(-y_i\alpha_t h_t(x_i)) \left(= 2\sqrt{\varepsilon_t(1-\varepsilon_t)}\right)$

  $D_{t+1}(i) \leftarrow \frac{1}{Z_t} D_t(i)\exp(-y_i\alpha_t h_t(x_i))$

end for

return a classifier F, $F(x) = \operatorname{sgn}\left(\sum\limits_{t=1}^{T}\alpha_t h_t(x)\right) = \operatorname{sgn}(f(x))$

我们陈述以下命题:

1. $\alpha_t = \arg\min\limits_{\alpha} Z_t = \arg\min\limits_{\alpha} \sum\limits_{i=1}^{n} D_t(i)\exp(-\alpha y_i h_t(x_i))$.

2. $\prod\limits_{t=1}^{T} Z_t = \frac{1}{n}\sum\limits_{i=1}^{n}\exp\left(-y_i\sum\limits_{t=1}^{T}\alpha_t h_t(x_i)\right) = \frac{1}{n}\sum\limits_{i=1}^{n}\exp(-y_i f(x_i))$.

3. $\sum\limits_{i=1}^{n} D_{t+1}(i)\mathbb{I}[y_i \neq h_t(x_i)] = \frac{1}{2}$. (与证明最终结果无关, 在此是为了指明每一轮中训练得到的 h_t 的优化方向是 orthogonal 的.)

4. $\mathbb{P}_{(x_i,y_i)\sim S}[F(x_i) \neq y_i] \leqslant (1-\gamma^2)^{T/2}$.

注意到 $\mathbb{P}_{(x_i,y_i)\sim S}[F(x_i) \neq y_i]$ 的最小非零结果应为 $\frac{1}{n}$, 故我们只需要限制 $(1-\gamma^2)^{T/2} < \frac{1}{n}$, 即 $T = \Omega(\log n)$, 就能将 error 降为 0.

1. Recall that $\varepsilon_t = \sum\limits_{i=1}^{n} D_t(i)\mathbb{I}[y_i \neq h_t(x_i)]$, by applying **AM-GM inequality** we have

$$Z_t = \sum_{i=1}^{n} D_t(i)\exp(-\alpha y_i h_t(x_i)) = \varepsilon_t\exp(\alpha) + (1-\varepsilon)\exp(-\alpha) \geqslant 2\sqrt{\varepsilon_t(1-\varepsilon_t)}$$

   where the equality holds if and only if

$$\varepsilon_t \mathrm{e}^{\alpha} = (1-\varepsilon_t)\mathrm{e}^{-\alpha} \Leftrightarrow \alpha = \frac{1}{2}\ln\frac{1-\varepsilon_t}{\varepsilon_t} = \frac{1}{2}\ln\frac{1+\gamma_t}{1-\gamma_t} = \alpha_t$$

   where $\gamma_t = 1 - 2\varepsilon$ in the assignment of α_t. This suggests that $\alpha_t = \arg\min\limits_{\alpha} Z_t$ as desired.

2. Notice that $D_{t+1}(i) = \frac{D_t(i)\exp(-\alpha_t y_i h_t(x_i))}{Z_t}$, by iteratively substitute the term of $D_t(i)$ in the expression

of Z_T ($Z_T = \sum_{i=1}^n D_T(i)\exp(-\alpha_T y_i h_T(x_i))$), we can eventually obtain the following equality.

$$\begin{aligned}
\prod_{t=1}^{T} Z_t &= \prod_{t=1}^{T-1} Z_t \sum_{i=1}^{n} D_T(i)\exp(-\alpha_T y_i h_T(x_i)) \\
&= \prod_{t=1}^{T-2} Z_t \sum_{i=1}^{n} D_{T-1}(i)\exp(-\alpha_T y_i h_T(x_i) - \alpha_{T-1} y_i h_{T-1}(x_i)) \\
&= \cdots \\
&= \sum_{i=1}^{n} D_0(i)\exp\left(-y_i \sum_{t=1}^{T} \alpha_t h_t(x_i)\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} \exp\left(-y_i f(x_i)\right)
\end{aligned}$$

3.

$$\begin{aligned}
\sum_{i=1}^{n} D_{t+1}(i)\mathbb{I}[y_i \neq h_t(x_i)] &= \sum_{i=1}^{n} \frac{D_t(i)\exp(-\alpha_t y_i h_t(x_i))}{Z_t}\mathbb{I}[y_i \neq h_t(x_i)] \\
&= \frac{\sum\limits_{i=1}^{n} D_t(i)\exp(\alpha_t)\mathbb{I}[y_i \neq h_t(x_i)]}{\sum\limits_{i=1}^{n} D_t(i)\exp(\alpha_t)\mathbb{I}[y_i \neq h_t(x_i)] + \sum\limits_{i=1}^{n} D_t(i)\exp(-\alpha_t)\mathbb{I}[y_i = h_t(x_i)]} \\
&= \frac{\varepsilon_t \mathrm{e}^{\alpha_t}}{\varepsilon_t \mathrm{e}^{\alpha_t} + (1-\varepsilon_t)\mathrm{e}^{-\alpha_t}}
\end{aligned}$$

Since α_t is chosen so that $\varepsilon_t \mathrm{e}^{\alpha_t} = (1-\varepsilon_t)\mathrm{e}^{-\alpha_t}$, we can prove that $\sum\limits_{i=1}^{n} D_{t+1}(i)\mathbb{I}[y_i \neq h_t(x_i)] = \frac{1}{2}$.

4.

$$\begin{aligned}
\mathbb{P}_{(x_i,y_i)\sim S}[F(x_i) \neq y_i] &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}[y_i f(x_i) \leqslant 0] \leqslant \frac{1}{n}\sum_{i=1}^{n} \exp(-y_i f(x_i)) = \prod_{t=1}^{T} Z_t \\
&= \prod_{t=1}^{T} 2\sqrt{\varepsilon_t(1-\varepsilon_t)} = \prod_{t=1}^{T} \sqrt{1-\gamma_t^2} \leqslant (1-\gamma^2)^{T/2}
\end{aligned}$$

6 PAC-Bayesian Theory

定理 6.1 (PAC-Bayesian Theorem). 对于给定的 prior distribution of classifiers $\mathcal{P}$, 从数据集 D 中随机抽取大小为 n 的训练集 S, 有至少 $1-\delta$ 的概率, 对于任意 distribution of classifiers $\mathcal{Q}$ 有如下不等式成立

$$\mathbb{E}_{h\sim\mathcal{Q}}[err_D(h)] \leqslant \mathbb{E}_{h\sim\mathcal{Q}}[err_S(h)] + \sqrt{\frac{D_{KL}(\mathcal{Q}\|\mathcal{P}) + \log(3/\delta)}{n}}$$

其中 $err_X(f)$ 表示 classifier f 在数据集 X 上的错误率, 即 $\mathbb{P}_{(x,y)\in X}[y \neq f(x)]$, $D_{KL}(\mathcal{Q}\|\mathcal{P}) = \mathbb{E}_{h\sim\mathcal{Q}}\left[\ln\frac{\mathcal{Q}_h}{\mathcal{P}_h}\right]$ 为概率分布 $\mathcal{Q}$ 与 $\mathcal{P}$ 的 KL 散度.

引理 6.2. 对于任意在 hypothesis space $\mathcal{F}$ 上的概率分布 $\mathcal{P},\mathcal{Q}$, 以及任意函数 $f:\mathcal{F}\to\mathbb{R}$, 都有

$$\mathbb{E}_{h\sim Q}[f(h)] \leqslant \ln\mathbb{E}_{h'\sim P}[\exp(f(h'))] + D_{KL}(\mathcal{Q}\|\mathcal{P})$$

证明.

$$\begin{aligned}
\text{RHS} - \text{LHS} &= \ln\mathbb{E}_{h'\sim P}[\exp(f(h'))] + D_{KL}(\mathcal{Q}\|\mathcal{P}) - \mathbb{E}_{h\sim Q}[f(h)] \\
&= \ln\mathbb{E}_{h'\sim P}[\exp(f(h'))] + \mathbb{E}_{h\sim\mathcal{Q}}\left[\ln\frac{\mathcal{Q}_h}{\mathcal{P}_h}\right] - \mathbb{E}_{h\sim Q}[f(h)] \\
&= \mathbb{E}_{h\sim Q}\left[\ln\frac{\mathcal{Q}_h}{\frac{\mathcal{P}_h\exp(f(h))}{\mathbb{E}_{h'\sim\mathcal{P}}[\exp(f(h'))]}}\right] \\
&= \mathbb{E}_{h\sim Q}\left[\ln\frac{\mathcal{Q}_h}{\mathcal{R}_h}\right] \\
&= D_{KL}(\mathcal{Q}\|\mathcal{R}) \\
&\geqslant 0
\end{aligned}$$

其中 $\mathcal{R}$ 也是一个 $\mathcal{F}$ 上的概率分布, $\mathcal{R}_h = \frac{\mathcal{P}_h\exp(f(h))}{\mathbb{E}_{h'\sim\mathcal{P}}[\exp(f(h'))]}$. □

引理 6.3. 对于任意 $\delta>0$, 有

$$\mathbb{P}_{S\sim D^n}\left(\mathbb{E}_{h\sim\mathcal{P}}[\mathrm{e}^{n(err_D(h)-err_S(h))^2}] \geqslant 3/\delta\right) \leqslant \delta$$

证明. 先证明对于某个固定的 $h\sim\mathcal{P}$, 有

$$\mathbb{E}_{S\sim D^n}[\mathrm{e}^{n(err_D(h)-err_S(h))^2}] \leqslant 3$$

记 $\Delta = |err_D(h) - err_S(h)|$, 根据 Chernoff bound, 有

$$\mathbb{P}_{S\sim D^n}(\Delta\geqslant\varepsilon) \leqslant 2\exp(-2n\varepsilon^2)$$

于是

$$\begin{aligned}
\mathbb{E}_{S\sim D^n}\left[\mathrm{e}^{n\Delta^2}\right] &= \int_0^{+\infty}\mathbb{P}_{S\sim D^n}\left(\mathrm{e}^{n\Delta^2}\geqslant t\right)\mathrm{d}t \\
&= \int_1^{+\infty}\mathbb{P}_{S\sim D^n}\left(\Delta\geqslant\sqrt{\frac{\ln t}{n}}\right)\mathrm{d}t + 1 \\
&\leqslant \int_1^{+\infty} 2\mathrm{e}^{-2\ln t}\mathrm{d}t + 1 \\
&= 3
\end{aligned}$$

随后, 使用 Markov Inequality 得到

$$\mathbb{P}_{S\sim D^n}\left(\mathbb{E}_{h\sim\mathcal{P}}[\mathrm{e}^{n\Delta^2}]\geqslant 3/\delta\right)\leqslant\frac{\mathbb{E}_{S\sim D^n}\left(\mathbb{E}_{h\sim\mathcal{P}}[\mathrm{e}^{n\Delta^2}]\right)}{3/\delta}=\frac{\mathbb{E}_{h\sim\mathcal{P}}\left(\mathbb{E}_{S\sim D^n}[\mathrm{e}^{n\Delta^2}]\right)}{3/\delta}\leqslant\frac{\mathbb{E}_{h\sim\mathcal{P}}(3)}{3/\delta}=\delta$$

□

我们利用上述两个引理证明定理 6.1. 有至少 $1-\delta$ 的概率,

$$\begin{aligned}
\left(\mathbb{E}_{h\sim\mathcal{Q}}[err_D(h)-err_S(h)]\right)^2 &\leqslant \mathbb{E}_{h\sim\mathcal{Q}}[\Delta^2]\\
&=\frac{1}{n}\mathbb{E}_{h\sim\mathcal{Q}}[n\Delta^2]\\
&\leqslant\frac{1}{n}\left(\ln\mathbb{E}_{h\sim P}\left[\mathrm{e}^{n\Delta^2}\right]+D_{KL}(\mathcal{Q}\|\mathcal{P})\right)\\
&\leqslant\frac{1}{n}\left(\ln(3/\delta)+D_{KL}(\mathcal{Q}\|\mathcal{P})\right)
\end{aligned}$$

其中第一行等号使用了 Cauchy Inequality, 第三行使用了引理 6.2 代入 $f(h)=n\Delta^2$, 第四行使用了引理 6.3, with probability at least $1-\delta$.

6.1 PAC-Bayesian Bound for SVM

命题 6.4. 对于任意的 distribution of classifiers $\mathcal{Q}$, 令 $g_\mathcal{Q}$ 为一个确定性二分类器, $g_\mathcal{Q}(x)=\mathrm{sgn}(\mathbb{E}_{h\sim\mathcal{Q}}h(x))$, 则

$$err_D(g_\mathcal{Q})\leqslant 2\mathbb{E}_{h\sim\mathcal{Q}}[err_D(h)]$$

证明. 如果 $g_\mathcal{Q}$ 在一个数据点 x 上出错, 则说明 $\mathcal{Q}$ 中至少一半的 classifier 都在 x 上出错. □

考虑两个 distribution of classifiers $\mathcal{P}=\mathcal{N}(\mathbf{0},I_d),\mathcal{Q}=\mathcal{N}(\mu\mathbf{w},I_d)$, 其中 $\|\mathbf{w}\|_2=1$, μ 是缩放系数. 此时 $g_\mathcal{Q}$ 就是传统理解下的 linear classifier $\mathbf{w}$ (这里不考虑常数 b).

根据定理 6.1 的结论, 我们有

$$err_D(g_\mathcal{Q})\leqslant 2\left[\mathbb{E}_{h\sim\mathcal{Q}}err_S(h)+\sqrt{\frac{D_{KL}(\mathcal{Q}\|\mathcal{P})+\log(3/\delta)}{n}}\right]$$

$$\begin{aligned}
D_{KL}(\mathcal{Q}\|\mathcal{P})&=\int_{\mathbb{R}^d}\frac{1}{(2\pi)^{d/2}}\exp\left[-\frac{1}{2}\|\mathbf{x}-\mu\mathbf{w}\|^2\right]\frac{1}{2}\left(\|\mathbf{x}\|^2-\|\mathbf{x}-\mu\mathbf{w}\|^2\right)\mathrm{d}\mathbf{x}\\
&=\int_\lambda\int_{\mathbf{y}\in\mathbb{R}^{d-1},\mathbf{y}\perp\mathbf{w}}\frac{1}{(2\pi)^{d/2}}\exp\left[-\frac{1}{2}\|\lambda\mathbf{w}+\mathbf{y}-\mu\mathbf{w}\|^2\right]\frac{1}{2}\left(\|\lambda\mathbf{w}+\mathbf{y}\|^2-\|\lambda\mathbf{w}+\mathbf{y}-\mu\mathbf{w}\|^2\right)\mathrm{d}\lambda\mathrm{d}\mathbf{y}\\
&=\int_\lambda\int_{\mathbf{y}\in\mathbb{R}^{d-1},\mathbf{y}\perp\mathbf{w}}\frac{1}{(2\pi)^{d/2}}\exp\left[-\frac{1}{2}(\lambda-\mu)^2-\frac{1}{2}\|\mathbf{y}\|^2\right]\frac{1}{2}\left(\lambda^2+\|\mathbf{y}\|^2-(\lambda-\mu)^2-\|\mathbf{y}\|^2\right)\mathrm{d}\lambda\mathrm{d}\mathbf{y}\\
&=\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}(\lambda-\mu)^2\right]\frac{1}{2}(2\lambda\mu-\mu^2)\mathrm{d}\lambda\left[\int_{\mathbf{y}\in\mathbb{R}^{d-1},\mathbf{y}\perp\mathbf{w}}\frac{1}{(2\pi)^{(d-1)/2}}\exp\left(-\frac{1}{2}\|\mathbf{y}\|^2\right)\mathrm{d}\mathbf{y}\right]\\
&=\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}(\lambda-\mu)^2\right](\lambda\mu-\mu^2)\mathrm{d}\lambda+\frac{\mu^2}{2}\\
&=\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}(\lambda-\mu)^2\right]\mu\mathrm{d}\frac{(\lambda-\mu)^2}{2}+\frac{\mu^2}{2}\\
&=\frac{\mu^2}{2}
\end{aligned}$$

7 Algorithmic Stability

定义 7.1 (**一致稳定, Uniform Stability**). $\mathcal{A}$ 是输入训练集 $S=(z_1,\cdots,z_n)$, 输出一个分类器 $\mathcal{A}(S)$ 的学习算法. 记 $S^i=(z_1,\cdots,z_{i-1},z_i',z_{i+1},\cdots,z_n)$ 是与 S 只相差第 i 个数据点的相邻训练集, $\ell(\cdot,\cdot)$ 是损失函数, 即 $\ell(f,z)$ 是在分类器 f 下, 数据点 z 产生的损失.

称学习算法 $\mathcal{A}$ 关于 $\ell(\cdot,\cdot)$ 满足 $\beta(n)$-一致稳定性, 如果对于任意大小为 n 的训练集 S 及其相邻训练集 S^i, 以及任意数据点 z, 都有

$$|\ell(\mathcal{A}(S),z)-\ell(\mathcal{A}(S^i),z)|\leqslant\beta(n)$$

定义 7.2 (**Risk & Empirical Risk**). 分别类似于 test error 与 training error, 定义 risk 与 empirical risk 为

$$R(\mathcal{A}(S))=\mathbb{E}_{z\sim D}[\ell(\mathcal{A}(S),z)]$$

$$R_{emp}(\mathcal{A}(S))=\frac{1}{n}\sum_{i=1}^{n}\ell(\mathcal{A}(S),z_i)$$

以下讨论中不会出现超过一个学习算法, 故简记 $\Phi(S)=R(\mathcal{A}(S))-R_{emp}(\mathcal{A}(S))$.

定理 7.3 (**一致稳定能说明泛化**). 对于一个关于 $\ell(\cdot,\cdot)$ 满足 $\beta(n)$-一致稳定性的学习算法 $\mathcal{A}$, 其中 $|\ell(\cdot,\cdot)|\leqslant M$ 有上界, 有

$$\mathbb{P}\left[\Phi(S)\leqslant\varepsilon+\beta(n)\right]\leqslant\exp\left(-\frac{n\varepsilon^2}{2(n\beta(n)+M)^2}\right)$$

或者等价的, 有至少 $1-\delta$ 的概率下式成立

$$R(\mathcal{A}(s))\leqslant R_{emp}(\mathcal{A}(s))+\beta(n)+(n\beta(n)+M)\sqrt{\frac{2\ln(1/\delta)}{n}}$$

证明. 先证明两个引理.

引理 7.4. 假设 $\mathcal{A}$ 是对称的, 即对于任意 n 元置换 σ, 有 $\mathcal{A}(\{z_1,\cdots,z_n\})=\mathcal{A}(\{z_{\sigma_1},\cdots,z_{\sigma_n}\})$, 则

$$\mathbb{E}_S[\Phi(S)]\leqslant\beta(n)$$

证明.

$$\mathbb{E}_S[\Phi(S)]=\mathbb{E}_{S,z}[\ell(\mathcal{A}(S),z)]-\mathbb{E}_S[\ell(\mathcal{A}(S),z_1)]=\mathbb{E}_{S,S^1}[\ell(\mathcal{A}(S^1),z_1)-\ell(\mathcal{A}(S),z_1)]\leqslant\beta(n)$$

□

引理 7.5. 如果 $|\ell(\cdot,\cdot)|\leqslant M$ 有上界, 则对于任意 S,S^i, 有

$$|\Phi(S)-\Phi(S^i)|\leqslant 2\left(\beta(n)+\frac{M}{n}\right)$$

证明. 除了 $\ell(\mathcal{A}(S),z_i)-\ell(\mathcal{A}(S^i),z_i')$ 一项外, 其余所有项都可以被 $\beta(n)$-稳定性限制住.

$$\begin{aligned}|\Phi(S)-\Phi(S^i)|&=|R(\mathcal{A}(S))-R_{emp}(\mathcal{A}(S))-R(\mathcal{A}(S^i))+R_{emp}(\mathcal{A}(S^i))|\\&\leqslant|R_{emp}(\mathcal{A}(S))-R_{emp}(\mathcal{A}(S^i))|+|R(\mathcal{A}(S))-R(\mathcal{A}(S^i))|\\&=\frac{1}{n}|\ell(\mathcal{A}(S),z_i)-\ell(\mathcal{A}(S^i),z_i')|+\frac{1}{n}\sum_{j\neq i}|\ell(\mathcal{A}(S),z_j)-\ell(\mathcal{A}(S^i),z_j)|+\left|\mathbb{E}_{z\sim D}[\ell(\mathcal{A}(S),z)-\ell(\mathcal{A}(S^i),z)]\right|\\&\leqslant\frac{2M}{n}+\frac{n-1}{n}\beta(n)+\beta(n)\\&\leqslant 2\left(\beta(n)+\frac{M}{n}\right)\end{aligned}$$

□

考虑 McDiarmid Inequality (定理 1.10), 把 Φ 视作一个关于 $z_1, \cdots, z_n$ 的多元函数, 则引理 7.4 与引理 7.5 分别给出了 Φ 的期望以及在相邻输入上的差的上界. 于是

$$\mathbb{P}\left[\Phi(S) \geqslant \beta(n) + \varepsilon\right] \leqslant \mathbb{P}\left[\Phi(S) - \mathbb{E}\left[\Phi(S)\right] \geqslant \varepsilon\right] \leqslant \exp\left(-\frac{2n\varepsilon^2}{\sum_{i=1}^n c_i^2}\right) = \exp\left(-\frac{n\varepsilon^2}{2(n\beta(n)+M)^2}\right)$$

□

8 Unsupervised Learning

前面讨论的都是监督学习. 现在我们讨论一下无监督学习.

无监督学习其实主要在做两件事情: Clustering, 以及 Dimensionality Reduction.

8.1 Clustering

对于一组 $\mathbf{x}_1, \cdots, \mathbf{x}_n \in \mathbb{R}^d$, 需要把这些数据点划分成 k 个 cluster $S_1, \cdots, S_k$.

可以如下定义一种划分的损失函数: 记 $\mu_i = \frac{1}{|S_i|}\sum_{j\in S_i}\mathbf{x}_j$ 为第 i 个 cluster 的中心, 损失函数为

$$L(\{S_1, \cdots, S_k\}) = \sum_{i=1}^{k}\sum_{j\in S_i}\|\mathbf{x}_j - \mu_i\|^2$$

8.1.1 K-means

Algorithm 2 K-means

1: Choose k points as cluster centers $\mu_1, \cdots, \mu_k$ uniformly at random.
2: **repeat**
3: $\quad S_i \leftarrow \{j : \|\mathbf{x}_j - \mu_i\|^2 \leqslant \|\mathbf{x}_j - \mu_k\|^2, \forall k \in [m]\}$
4: $\quad \mu_i \leftarrow \frac{1}{|S_i|}\sum_{j\in S_i}\mathbf{x}_j$
5: **until** k cluster centers do not change
6: **return** $\{\mu_1, \cdots, \mu_k\}$

8.1.2 K-means++

Algorithm 3 K-means++

1: Choose a point as the cluster center μ_1 uniformly at random.
2: **for** $i : 2 \to n$ **do**
3: $\quad$ Choose a point as the cluster center μ_i, with probability proportional to $\min\limits_{1\leqslant k<i}\|\mathbf{x}_j - \mu_k\|^2$.
4: **end for**
5: **return** $\{\mu_1, \cdots, \mu_k\}$

定理 8.1. K-means++ 算法给出的损失 L 与最优解 L_{opt} 满足

$$\mathbb{E}[L] \leqslant 8(\ln k + 2)L_{opt}$$

8.2 Dimensionality Reduction

wlw 不讲了.

9 Online Learning

在在线学习的设定下, 数据是以流的形式给出的, 在每次得到一个数据点之后, 都需要以恰当的方式更新预测器, 以优化将来的预测.

相比监督学习, 在线学习主要区别在于: (1) 不再区分 training 与 test, (2) 没有对数据的分布假设, 因而不存在 generalization 的概念. 相应的, mistake model 以及 regret 的概念会被用于衡量在线学习算法的表现效果.

9.1 Online Learning with Expert Advice

有 n 位专家. 预测会持续 T 轮, 每轮中每位专家都会给出各自的预测 $\tilde{y_{t,i}} \in \{0,1\}$, 学习者需要根据此前得到的所有信息给出预测 $\tilde{y_t} \in \{0,1\}$, 同时也会获得正确结果 $y_t \in \{0,1\}$. 学习者的目标是让自己的预测结果与最好的专家尽量接近, 即最小化 $\sum\limits_{t=1}^{T} \mathbb{1}[\tilde{y_t} \neq y_t]$ 与 $\min\limits_{i\in[n]} \sum\limits_{t=1}^{T} \mathbb{1}[\tilde{y_{t,i}} \neq y_t]$ 的差 (这就是 regret).

9.1.1 Weighted Majority Vote

Algorithm 4 Weighted Majority Vote

1: Initialize $w_{1,i} \leftarrow 1, \forall i \in [n]$
2: Choose parameter $\beta \in (0,1)$
3: **for** $t = 1 \to T$ **do**
4: Make the Weighted Majority Vote $\tilde{y_t} = \begin{cases} 0, & \sum\limits_{\tilde{y_{t,i}}=0} > \sum\limits_{\tilde{y_{t,i}}=1} \\ 1, & \text{otherwise} \end{cases}$
5: **if** $\tilde{y_t} = y_t$ **then**
6: $w_{t+1,i} \leftarrow w_{t,i}, \forall i \in [n]$
7: **else**
8: $w_{t+1,i} \leftarrow \begin{cases} \beta \cdot w_{t,i}, & \tilde{y_{t,i}} \neq y_t \\ w_{t,i}, & \tilde{y_{t,i}} = y_t \end{cases}, \forall i \in [n]$
9: **end if**
10: **end for**

即每轮选择 $\tilde{y_t}$ 为 n 位专家预测的加权 marjority, 如果出错了, 就把所有导致自己出错的专家的权值乘上 β 作为惩罚.

定理 9.1. 记 $L_T = \sum\limits_{t=1}^{T} \mathbb{1}[\tilde{y_t} \neq y_t]$ 为学习者的 loss, $m_T^* = \min\limits_{i\in[n]} \sum\limits_{t=1}^{T} \mathbb{1}[\tilde{y_{t,i}} \neq y_t]$ 为最好的专家的 loss, 则在 Weighted Majority Vote 算法下, 有

$$L_T \leqslant \frac{m_T^* \log(1/\beta) + \log n}{\log(2/(1+\beta))}$$

证明. 注意到 (1) T 轮结束后, 所有专家剩余的总权值至少还有 $\beta^{m_T^*}$, (2) 每次学习者出错都会导致总权值乘上不大于 $\frac{1+\beta}{2}$ 的系数, 故

$$\beta^{m_T^*} \leqslant n \left(\frac{1+\beta}{2}\right)^{L_T} \Rightarrow L_T \leqslant \frac{m_T^* \log(1/\beta) + \log n}{\log(2/(1+\beta))}$$

□

注 9.2. 考虑 $\beta \to 1$, 由 L'Hospital Rule 可知 $\frac{\log(1/\beta)}{\log(2/(1+\beta))} \to 2$, 即 Weighted Majority Vote 算法给出的最好的界中, m_T^* 前的系数至少是 2. 接下来的 Randomized Weighted Updating 算法会给出更好的界.

Algorithm 5 Randomized Weighted Updating

Initialize $w_{1,i} \leftarrow 1, \forall i \in [n]$
Choose parameter $\beta \in [\frac{1}{2}, 1)$
for $t = 1 \to T$ **do**
  Chooes $\tilde{y_t} = \tilde{y_{t,i}}$ with probability proportional to $w_{t,i}$
  $w_{t+1,i} \leftarrow \begin{cases} \beta \cdot w_{t,i}, & \tilde{y_{t,i}} \neq y_t \\ w_{t,i}, & \tilde{y_{t,i}} = y_t \end{cases}, \forall i \in [n]$
end for

9.1.2 Randomized Weighted Updating

定理 9.3. 在 Randomized Weighted Updating 算法下, 有

$$\mathbb{E}\left[L_T\right] \leqslant (2-\beta) m_T^* + \frac{\ln n}{1-\beta}$$

证明. 注意到权值的更新无关于每轮有没有答错, 因此 $\mathbb{1}[\tilde{y_t} \neq y_t]$ 是独立随机变量.

第 i 轮结束后, 总权值的变化一定是 $W \to W(1-(1-\beta)\mathbb{P}\left[\tilde{y_t} \neq y_t\right])$, 由于 $\mathbb{E}\left[L_T\right] = \sum_{t=1}^{T} \mathbb{P}\left[\tilde{y_t} \neq y_t\right]$, 因此

$$\beta^{m_T^*} \leqslant n \prod_{t=1}^{T}(1-(1-\beta)\mathbb{P}\left[\tilde{y_t} \neq y_t\right]) \leqslant n \prod_{t=1}^{T} \mathrm{e}^{-(1-\beta)\mathbb{P}\left[\tilde{y_t} \neq y_t\right]} = n\mathrm{e}^{-(1-\beta)\mathbb{E}\left[L_T\right]}$$

从而得到了

$$\mathbb{E}\left[L_T\right] \leqslant \frac{\ln(1/\beta) m_T^* + \ln n}{1-\beta}$$

只需要进一步证明 $\frac{\ln(1/\beta)}{1-\beta} \leqslant 2-\beta$. 考虑函数 $f(\beta) = \ln\beta + (1-\beta)(2-\beta)$, $f'(\beta) = \frac{(1-\beta)(1-2\beta)}{\beta}$, 当 $\beta \in [\frac{1}{2}, 1)$ 时恒有 $f'(\beta) \leqslant 0$, 从而 $f(\beta) \geqslant f(1) = 0$, 说明了 $\ln(1/\beta) \leqslant (1-\beta)(2-\beta)$, $\frac{\ln(1/\beta)}{1-\beta} \leqslant 2-\beta$. □

9.1.3 Hedge Algorithm

我们再提出一种叫做 Hedge Algorithm 的算法, 它其实只是 Randomized Weighted Updating 的推广, 但这个结果可以为后续证明定理 3.3 的工作做准备.

在 Hedge Algorithm 的设定下, loss 不再是"答错了几次", 而是每一轮每一位专家的回答都有一个 loss $g_t(i) \in [0,1]$, 记学习者在第 t 轮的 loss 为 l_t, 则 l_t 的期望就是 n 位专家的加权平均:

$$\mathbb{E}\left[l_t\right] = \left(\sum_{i=1}^{n} w_{t,i} g_t(i)\right) \Big/ \left(\sum_{i=1}^{n} w_{t,i}\right)$$

Algorithm 6 Hedge Algorithm

Initialize $w_{1,i} \leftarrow 1, \forall i \in [n]$
Choose parameter $\beta \in (0,1)$
for $t = 1 \to T$ **do**
  Chooes $i_t \in [n]$ with probability proportional to $w_{t,i}$, and obtain the loss $l_t = g_t(i_t)$
  $w_{t+1,i} \leftarrow w_{t,i} \cdot \beta^{g_t(i)}, \forall i \in [n]$
end for

定理 9.4. 重新定义 $L_T = \sum_{t=1}^{T} l_t$, 在 Hedge Algorithm 下, 有

$$\mathbb{E}\left[L_T\right] - \min_{i\in[n]} \sum_{t=1}^{T} g_t(i) = O(\sqrt{T\log n})$$

证明. 仍然注意到 l_t 是独立随机变量.

第 i 轮结束后, 总权值的变化是 $W \to W \cdot \mathbb{E}\left[\beta^{l_t}\right]$, 从而有

$$\begin{aligned}
\mathrm{e}^{-\ln(1/\beta)m_T^*} = \beta^{m_T^*} &\leqslant n\prod_{t=1}^{T}\mathbb{E}\left[\beta^{l_t}\right] = n\prod_{t=1}^{T}\mathbb{E}\left[\mathrm{e}^{-\ln(1/\beta)l_t}\right] \\
&\leqslant n\prod_{t=1}^{T}\mathbb{E}\left[1-\ln(1/\beta)l_t+\ln^2(1/\beta)l_t^2\right] \\
&\leqslant n\prod_{t=1}^{T}\left(1-\ln(1/\beta)\mathbb{E}\left[l_t\right]+\ln^2(1/\beta)\right) \\
&\leqslant n\prod_{t=1}^{T}\mathrm{e}^{-\ln(1/\beta)\mathbb{E}[l_t]+\ln^2(1/\beta)} \\
&= n\mathrm{e}^{-\ln(1/\beta)\mathbb{E}[L_T]+T\ln^2(1/\beta)}
\end{aligned}$$

其中 $m_T^* = \min_{i\in[n]}\sum_{t=1}^{T} g_t(i)$. 两边取对数得到

$$\mathbb{E}\left[L_T\right] - \min_{i\in[n]} \sum_{t=1}^{T} g_t(i) \leqslant \frac{\ln n}{\ln(1/\beta)} + T\ln(1/\beta) \leqslant 2\sqrt{T\ln n} = O(\sqrt{T\log n})$$

□

9.2 Proof of Minimax Theorem via Online Learning

在 Game Theory 一章中, 我们陈述了 Minimax Theorem (定理 3.3), 其表明在混合策略的双人零和博弈下, 先后手并不会影响博弈的最终结果. 接下来我们利用在线学习的技术来证明这个结论.

$$\min_p \max_q p^{\mathrm{T}}Mq = \max_q \min_p p^{\mathrm{T}}Mq$$

9.2.1 The $\geqslant$ Direction

这个方向的结论应该是平凡的, 直观上来说就是 "后手总不劣于先手".

形式化地, 记 $p^* = \arg\min_p \max_q p^{\mathrm{T}}Mq$ 为 row player 后手时选择的最优的 p, $q^* = \arg\max_q \min_p p^{\mathrm{T}}Mq$ 为 column player 后手时选择的最优的 q, 则

$$\min_p \max_q p^{\mathrm{T}}Mq = \max_q {p^*}^{\mathrm{T}}Mq \geqslant {p^*}^{\mathrm{T}}Mq^* \geqslant \min_p p^{\mathrm{T}}Mq^* = \max_q \min_p p^{\mathrm{T}}Mq$$

9.2.2 The $\leqslant$ Direction

row player 对应在线学习中的学习者, column player 对应 adversary, 收益矩阵 M 的 m 行分别是一位专家.

在第 t 轮中, 学习者选择列向量 p_t 满足 $(p_t)_i = \frac{w_{t,i}}{\sum_{i=1}^{m} w_{t,i}}$, 其中 $w_{t,i}$ 表示第 t 轮时第 i 位专家的权值. 给出了 p_t 后, adversary 可以很容易地给出 $q_t = \max_q p_t^{\mathrm{T}}Mq$. 第 i 位专家建议选第 i 行, 他这样的方案对应的 loss 是 $g_t(i) = (Mq_t)_i$. 显然学习者此时的 loss 的期望恰好等于 m 为专家各自损失的加权平均, 即

$$\mathbb{E}\left[l_t\right] = \left(\sum_{i=1}^{n} w_{t,i}g_t(i)\right) \Bigg/ \left(\sum_{i=1}^{n} w_{t,i}\right) = p_t^{\mathrm{T}}Mq_t$$

由 Hedge Algorithm 以及定理 9.4, 我们知道了

$$\mathbb{E}\left[L_T\right] - \min_{i\in[n]}\sum_{t=1}^{T} g_t(i) = \sum_{t=1}^{T} p_t^{\mathrm{T}} M q_t - \min_{i\in[n]}\left(M\sum_{t=1}^{T} q_t\right)_i \leqslant O(\sqrt{T\log m})$$

由此得到

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T} p_t^{\mathrm{T}} M q_t &\leqslant \min_{i\in[n]}\left(M\sum_{t=1}^{T} q_t\right)_i + O\left(\sqrt{\frac{\log m}{T}}\right)\\ &= \min_p\left(p^{\mathrm{T}} M\left(\frac{1}{T}\sum_{t=1}^{T} q_t\right)\right) + o(1)\\ &\leqslant \max_q\min_p p^{\mathrm{T}} M q + o(1)\end{aligned}$$

(其中 $O\left(\sqrt{\frac{\log m}{T}}\right) = o(1)$ 因为我们视 m 为常数) 而又注意到

$$\min_p\max_q p^{\mathrm{T}} M q \leqslant \max_q\left(\frac{1}{T}\sum_{t=1}^{T} p_t^{\mathrm{T}}\right) M q \leqslant \frac{1}{T}\sum_{t=1}^{T}\max_q p_t^{\mathrm{T}} M q = \frac{1}{T}\sum_{t=1}^{T} p_t^{\mathrm{T}} M q_t$$

因此 $\min_p\max_q p^{\mathrm{T}} M q \leqslant \max_q\min_p p^{\mathrm{T}} M q + o(1)$, 即 $\min_p\max_q p^{\mathrm{T}} M q \leqslant \max_q\min_p p^{\mathrm{T}} M q$.

9.3 Multi-arm Bandits (MAB) Problem

有一台多臂老虎机, 它有 k 个拉杆, 第 i 个拉杆拉动后会返回一个服从分布 $\mathcal{D}_i$ 的随机变量 loss, 其中 $\mathcal{D}_i$ 的均值为 μ_i. 需要最小化拉 T 轮后得到的 loss 之和.

记第 t 轮中选择拉动了第 a_t 个拉杆, 我们可以定义 T 轮操作后的 regret 为

$$R_T = \mathbb{E}_{\mathcal{A}}\left[\sum_{t=1}^{T}\mu_{a_t} - \mu^*\right]$$

其中 $\mathcal{A} = (a_1, \cdots, a_T)$ 为选择拉动的拉杆编号序列, $\mu^* = \min_{1\leqslant i\leqslant k}\mu_i$ 为最优的期望 loss. 概率源自于 $\mathcal{A}$ 中 a_t 的选取会基于之前随机变量 $l_1 \sim \mathcal{D}_{a_1}, \cdots, l_{t-1} \sim \mathcal{D}_{t-1}$ 的实际取值.

由于在 MAB 问题中我们并不先验地知道每个分布的均值, 因此这是一个在 exploration (调查每个拉杆) 和 exploitation (对着“最好”的薅) 之间权衡的过程.

9.3.1 UCB Algorithm

Algorithm 7 UCB Algorithm

1: $n_t(a)$ represents # of times arm a has been pulled at time t
2: $\mu_t(a)$ represents the empirical loss of arm i at time t
3: Initialize $n_0(a) \leftarrow 0, \mu_0(a) \leftarrow 0$
4: **for** $t = 1 \to T$ **do**
5:     For each arm a, compute $\mathrm{UCB}_t(a) \leftarrow \mu_{t-1}(a) - \sqrt{\frac{\ln T}{n_{t-1}(a)}}$
6:     Pull the arm $a_t = \arg\min_{1\leqslant a\leqslant k}\mathrm{UCB}_t(a)$
7:     Update $n_t(a)$ and $\mu_t(a)$ for each $1 \leqslant a \leqslant k$
8: **end for**

特别地, 当 $n_{t-1}(a) = 0$ 时, 记 $\mathrm{UCB}_t(a) = -\infty$.

定理 9.5. 不失一般性假设 $\mu_1 \leqslant \min\{\mu_2, \cdots, \mu_k\}$, UCB Algorithm 得到的 regret 可以被限制为

$$R_T = \mathbb{E}_{\mathcal{A}}\left[\sum_{t=1}^{T} \mu_{a_t} - \mu_1\right] \leqslant \sum_{\Delta_a > 0}\left(\frac{16 \ln T}{\Delta_a} + 2\Delta_a\right)$$

其中 $\Delta_a = \mu_a - \mu_1$.

证明. 注意到

$$R_T = \mathbb{E}_{\mathcal{A}}\left[\sum_{t=1}^{T} \mu_{a_t} - \mu_1\right] = \sum_{a=1}^{k} \Delta_a \cdot \mathbb{E}_{\mathcal{A}}[n_T(a)]$$

只需要证明对于每个 $\Delta_a > 0$ 的拉杆 a, 都有

$$\mathbb{E}_{\mathcal{A}}[n_T(a)] \leqslant \frac{16 \ln T}{\Delta_a^2} + 2$$

对于任意正整数 m, 我们有

$$\begin{aligned}
\mathbb{E}_{\mathcal{A}}[n_T(a)] &= \sum_{t=1}^{T} \mathbb{P}\left[a_t = a\right] \\
&= \sum_{t=1}^{T} \mathbb{P}\left[a_t = a \wedge n_{t-1}(a) < m\right] + \sum_{t=1}^{T} \mathbb{P}\left[a_t = a \wedge n_{t-1}(a) \geqslant m\right] \\
&\leqslant m + \sum_{t=m+1}^{T} \mathbb{P}\left[a_t = a \wedge n_{t-1}(a) \geqslant m\right]
\end{aligned}$$

命题 9.6. 假如某个 $\Delta_a > 0$ 的拉杆 a 在第 t 轮被拉动了, 则要么 $\mathrm{UCB}_t(1) > \mu_1$, 要么 $\mathrm{UCB}_t(a) < \mu_1 = \mu_a - \Delta_a$.

证明. 否则 $\mathrm{UCB}_t(1) \leqslant \mu_1 \leqslant \mathrm{UCB}_t(a)$, 拉动 a 不如拉动 1. □

利用上述命题,

$$\begin{aligned}
\mathbb{P}\left[a_t = a \wedge n_{t-1}(a) \geqslant m\right] &= \mathbb{P}\left[(\mathrm{UCB}_t(1) > \mu_1 \vee \mathrm{UCB}_t(a) < \mu_1) \wedge n_{t-1}(a) \geqslant m\right] \\
&\leqslant \mathbb{P}\left[\mathrm{UCB}_t(1) > \mu_1 \wedge n_{t-1}(a) \geqslant m\right] + \mathbb{P}\left[\mathrm{UCB}_t(a) < \mu_1 \wedge n_{t-1}(a) \geqslant m\right] \\
&\leqslant \mathbb{P}\left[\mathrm{UCB}_t(1) > \mu_1\right] + \mathbb{P}\left[\mathrm{UCB}_t(a) < \mu_1 \wedge n_{t-1}(a) \geqslant m\right]
\end{aligned}$$

对于前一部分,

$$\begin{aligned}
\mathbb{P}\left[\mathrm{UCB}_t(1) > \mu_1\right] &= \sum_{k=1}^{T} \mathbb{P}\left[\mu_{t-1}(1) - \sqrt{\frac{\ln T}{k}} > \mu_1 \wedge n_{t-1}(1) = k\right] \\
&\leqslant \sum_{k=1}^{T} \mathbb{P}\left[\mu_{t-1}(1) - \mu_1 > \sqrt{\frac{\ln T}{k}} \middle| n_{t-1}(1) = k\right] \\
&\leqslant \sum_{k=1}^{T} \exp\left(-2k\left(\sqrt{\frac{\ln T}{k}}\right)^2\right) \\
&= \sum_{k=1}^{T} \frac{1}{T^2} = \frac{1}{T}
\end{aligned}$$

其中第三行用到了 Chernoff Bound (定理 1.8).

对于第二个, 取 m 满足 $2\sqrt{\frac{\ln T}{m}} \leqslant \Delta_a \leqslant 4\sqrt{\frac{\ln T}{m}}$, 有 $m \leqslant \frac{16 \ln T}{\Delta_a^2}$,

$$\begin{aligned}
\mathbb{P}\left[\mathrm{UCB}_t(a) < \mu_1 \wedge n_{t-1}(a) \geqslant m\right] &= \mathbb{P}\left[\mu_{t-1}(a) - \mu_a < \sqrt{\frac{\ln T}{n_{t-1}(a)}} - \Delta_a \wedge n_{t-1}(a) \geqslant m\right] \\
&\leqslant \mathbb{P}\left[\mu_{t-1}(a) - \mu_a < -\sqrt{\frac{\ln T}{m}}\right] \leqslant \frac{1}{T}
\end{aligned}$$

结合上述结果, 我们得到

$$\mathbb{E}_{\mathcal{A}}[n_T(a)] \leqslant m + \sum_{t=m+1}^{T} \mathbb{P}\left[a_t = a \wedge n_{t-1}(a) \geqslant m\right] \leqslant m + \sum_{t=m+1}^{T} \frac{2}{T} \leqslant \frac{16\ln T}{\Delta_a^2} + 2$$

完成了证明. □

注 9.7. 上述结论是 instance-dependent 的, 因为其限制中包含了 Δ_a 项. 如果我们把 Δ_a 视作常数, 那么 R_T 就是 $O(k\ln T)$ 级别, 这比 Online Learning with Expert Advice 的 $O(\sqrt{T\log n})$ 要厉害. 不过, 如果 Δ_a 很小, 上述结论给出的界就会很差. 接下来我们给出一个更精细化的结论.

定理 9.8. 假设 Δ_a 有界 (存在 $M > 0$ 使得 $\Delta_a \leqslant M$ 成立), 则最坏情况下 UCB Algorithm 得到的 regret 为

$$R_T = O(\sqrt{kT\ln T})$$

证明. 取 $\delta = \sqrt{\frac{k\ln T}{T}}$, 将所有 a 按照 Δ_a 与 δ 的大小关系分为两组:

$$\begin{aligned}
R_T^{(1)} &= \sum_{t=1}^{T} \sum_{0<\Delta_a<\delta} \Delta_a \cdot \mathbb{P}\left[a_t = a\right] \leqslant T \cdot \delta \leqslant \sqrt{kT\ln T} \\
R_T^{(2)} &= \sum_{\Delta_a \geqslant \delta} \left(\frac{16\ln T}{\Delta_a} + 2\Delta_a\right) = O\left(\frac{kT}{\delta} + k\right) = O(\sqrt{kT\ln T}) \\
R_T &= R_T^{(1)} + R_T^{(2)} = O(\sqrt{kT\ln T})
\end{aligned}$$

□

9.3.2 Thompson Sampling

定义 9.9 (*Beta* **分布**)**.** *Beta* 分布是在区间 $(0,1)$ 上的连续分布, 对于参数 $\alpha, \beta > 0$, 分布 $Beta(\alpha, \beta)$ 的概率密度函数为

$$f(x;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

在 Thompson Sampling 中, 我们假设所有 loss 都是 $[0,1]$ 的.

Algorithm 8 Thompson Sampling

1: Initialize $S_a \leftarrow 0, F_a \leftarrow 0$
2: **for** $t = 1 \to T$ **do**
3: For each arm a, sample $\Theta_a(t) \sim Beta(S_a + 1, F_a + 1)$
4: Pull the arm $a_t = \arg\max_{1\leqslant a\leqslant k} \Theta_a(t)$, and obtain the loss $l_t \sim \mathcal{D}_{a_t}$ ▷ $l_t \in [0,1]$
5: Sample $\tilde{l}_t \sim \mathcal{B}(1, l_t)$ ▷ $\tilde{l}_t \in \{0,1\}$
6: $S_{a_t} \leftarrow S_{a_t} + \tilde{l}_t, F_{a_t} \leftarrow F_{a_t} + 1 - \tilde{l}_t$
7: **end for**

定理 9.10. 不失一般性假设 $\mu_1 \leqslant \min\{\mu_2, \cdots, \mu_k\}$, 对于任意 $\varepsilon > 0$, Thompson Sampling 得到的 regret 都满足

$$R_T \leqslant (1+\varepsilon) \sum_{\mu_a \neq \mu_1} \frac{\Delta_a \log T}{D(\mu_a \| \mu_1)} + O\left(\frac{k}{\varepsilon^2}\right) \leqslant (1+\varepsilon) \sum_{\mu_a \neq \mu_1} \frac{\log T}{2\Delta_a} + O\left(\frac{k}{\varepsilon^2}\right)$$

证明太长了, 就不讲了.

10 Differential Privacy

设计差分隐私的主要目的, 是在不透露过多个体信息 (privacy) 的前提下, 提供尽量多, 或者尽量准确的整体统计信息 (non-privacy). 有一种简单的想法, 就是给输出的整体信息加 noise.

定义 10.1 (相邻数据集). 考虑 $D=\{x_1,x_2,\cdots,x_n\}, x_i\in\mathcal{X}$ 是单个数据点, 则 $D\in\mathcal{X}^n$ 是大小为 n 的数据集. 两个数据集 D,D' 被称为相邻的, 如果其只存在一位不同.

定义 10.2 (统计查询). 一个依据 $h:\mathcal{X}\to\{0,1\}$ 定义的统计查询 Q 是 $\mathcal{X}^n\to\mathbb{Q}$ 的映射, 满足

$$Q(D)=\frac{1}{|D|}\sum_i h(x_i)$$

定义 10.3 (差分隐私). 令 A 为一个在输入数据集 D 上运行的随机算法, 称 A 满足 ε-差分隐私, 如果对于任意相邻数据集 $D,D'\in\mathcal{X}^n$, 任意 $S\subseteq \operatorname{im} A$, 都有

$$\mathbb{P}\left[A(D)\in S\right]\leqslant \mathrm{e}^{\varepsilon}\mathbb{P}\left[A(D')\in S\right]$$

定义 10.4 ((α,β)-精确). 称随机算法 A 对于统计查询 Q 满足 (α,β)-精确, 如果对于任意 $D\in\mathcal{X}^n$,

$$\mathbb{P}\left[|A(D)-Q(D)|\geqslant\alpha\right]\leqslant\beta$$

10.1 Laplace Mechanism

定义 10.5 (拉普拉斯分布). 随机变量 X 服从参数为 μ,σ 的拉普拉斯分布 (记作 $X\sim\mathrm{Lap}(\mu,\sigma)$), 其密度函数为

$$f_X(x)=\frac{1}{2\sigma}\exp\left(\frac{|x-\mu|}{\sigma}\right)$$

定义 10.6 (Laplace Mechanism, or Additive noise mechanism). Laplace Mechanism 就是在一个统计查询 $Q:\mathcal{X}^n\to\mathbb{Q}$ 的基础上, 添加一个服从分布 $\mathrm{Lap}(\mu=0,\sigma)$ 的 noise, 得到 $A:\mathcal{X}^n\to\mathbb{R}$, $A(D)=Q(D)+Z, Z\sim\mathrm{Lap}(0,\sigma)$.

定理 10.7. Laplace Mechanism 满足 ε-差分隐私以及 (α,β)-精确, 其中 (假设 β 是常数) $\varepsilon=\frac{1}{n\sigma},\alpha=\sigma\ln\frac{1}{\beta}$.

证明. 对于任意 $a\in\operatorname{im} A=\mathbb{R}$, 有

$$\frac{\mathbb{P}\left[A(D)=a\right]}{\mathbb{P}\left[A(D')=a\right]}=\frac{f_Z(a-Q(D))}{f_Z(a-Q(D'))}=\frac{\frac{1}{2\sigma}\exp\left(-\frac{a-Q(D)}{\sigma}\right)}{\frac{1}{2\sigma}\exp\left(-\frac{a-Q(D')}{\sigma}\right)}\leqslant\exp\left(\frac{|Q(D)-Q(D')|}{\sigma}\right)\leqslant\exp\left(\frac{1}{n\sigma}\right)$$

[1]故 $\varepsilon=\frac{1}{n\sigma}$. 至于 α,

$$\mathbb{P}\left[|A(D)-Q(D)|\geqslant\alpha\right]=\int_{-\infty}^{-a}\frac{1}{2\sigma}\exp\left(-\frac{t}{\sigma}\right)\mathrm{d}t+\int_a^{\infty}\frac{1}{2\sigma}\exp\left(-\frac{t}{\sigma}\right)\mathrm{d}t=\exp\left(-\frac{\alpha}{\sigma}\right)=\beta\Rightarrow\alpha=\sigma\ln\frac{1}{\beta}$$

□

定义 10.8 (多组查询下的精确). 记 $Q=(Q_1,\cdots,Q_k):\mathcal{X}^n\to\mathbb{Q}^k$ 是 k 次统计查询, 令 $A=(A_1,\cdots,A_k)$ 为针对每个查询, 用 Laplace Mechanism 构造的随机算法, $A_i(D)-Q_i(D)\sim$ i.i.d. $\mathrm{Lap}(0,\sigma)$. 称 A 对于 Q 满足 (α,β)-精确, 如果对于任意 $D\in\mathcal{X}^n$,

$$\mathbb{P}\left[\|A(D)-Q(D)\|_\infty\geqslant\alpha\right]\leqslant\beta$$

定理 10.9. 当每个 Laplace Mechanism A_i 都满足 ε-差分隐私和 (α,β)-精确时, $A=(A_1,\cdots,A_k)$ 满足 $k\varepsilon$-差分隐私和 $(\alpha,k\beta)$-精确.

[1]这里 $A(D)=a$ 实际上想表达的是 $a\leqslant A(D)<a+\mathrm{d}t$.

证明. • $k\varepsilon$-差分隐私: 注意到 noise 是 i.i.d. 的, 故联合分布密度等于各自相乘, 由 $f_{A_i(D)}(x_i) \leqslant \mathrm{e}^{\varepsilon} f_{A_i(D')}(x_i)$ 可以很容易得到 $f_{A_i(D)}(\vec{x}) \leqslant \mathrm{e}^{k\varepsilon} f_{A_i(D')}(\vec{x})$.

- $(\alpha, k\beta)$-精确: 使用 Union Bound 即可.

□

推论 10.10. 对于多组查询 $Q = (Q_1, \cdots, Q_k)$, Laplace Mechanism 满足 $\frac{k}{n\sigma}$-差分隐私以及 $(\sigma \ln \frac{k}{\beta}, \beta)$-精确.

注意到在多组查询的 Laplace Mechanism 下, 如果要求 $\varepsilon = O(1), \alpha = o(1)$, β 是常数, 则不得不有 $k = o\left(\frac{n}{\ln n}\right)$. 换句话说, 隐私泄露 ε 是关于查询次数 k 线性的, 从某种程度上说, 这是不能接受的.

接下来我们将介绍一种 highly-nontrivial 的机制设计, 使得可以在保证隐私与精确的前提下做到 $k >> n$ 的查询次数.

10.2 BLR Mechanism

不妨假设样本空间是有限大的, 即 $|\mathcal{X}| = N$. 此外沿用之前的一些记号, k 表示查询次数, $n = |D|$ 表示单个数据集大小, ε 表示隐私性, (α, β) 表示精确性. 额外记 $\sigma = \frac{2}{n\varepsilon}$.

该机制采用的随机算法 $\mathcal{A}$ 不再返回一个 $\mathbb{R}^d$ 向量, 取而代之的是返回 $\mathcal{X}^m$ 即 m 个样本, 其中 $m = \frac{2\log(2k)}{\alpha^2}$. 对于 $D \in \mathcal{X}^n$, $\mathcal{A}(D)$ 返回 $\hat{D} \in \mathcal{X}^m$ 的概率 $\mathbb{P}\left[\mathcal{A}(D) = \hat{D}\right]$ 正比于 $\exp\left(\frac{u(D,\hat{D})}{\sigma}\right)$, 其中 u 是**<u>效用函数 (utility function)</u>**, 用于衡量「在输入 D 时返回 $\hat{D}$ 有多好」, 在这里定义为

$$u(D, \hat{D}) = -\max_{i=1}^{k} |Q_i(D) - Q_i(\hat{D})|$$

其中 Q_i 表示第 i 个统计查询.

定理 10.11. BLR Mechanism 满足 ε-差分隐私以及 (α, β)-精确, 其中

$$\alpha = O\left(\left(\frac{\log k \log N + \log(1/\beta)}{n\varepsilon}\right)^{1/3}\right)$$

证明. 记 $\Delta u = \max_{D,D',\hat{D}} |u(D, \hat{D}) - u(D', \hat{D})|$, 可以观察到 $\Delta u \leqslant \frac{1}{n}$.

考虑

$$\mathbb{P}\left[\mathcal{A}(D) = \hat{D}\right] = \frac{\exp\left(\frac{u(D,\hat{D})}{\sigma}\right)}{\sum_{\hat{D}} \exp\left(\frac{u(D,\hat{D})}{\sigma}\right)}$$
$$\mathbb{P}\left[\mathcal{A}(D') = \hat{D}\right] = \frac{\exp\left(\frac{u(D',\hat{D})}{\sigma}\right)}{\sum_{\hat{D}} \exp\left(\frac{u(D',\hat{D})}{\sigma}\right)}$$

从而得到

$$\begin{aligned}
\frac{\mathbb{P}\left[\mathcal{A}(D) = \hat{D}\right]}{\mathbb{P}\left[\mathcal{A}(D') = \hat{D}\right]} &\leqslant \frac{\sum_{\hat{D}} \exp\left(\frac{u(D',\hat{D})}{\sigma}\right)}{\sum_{\hat{D}} \exp\left(\frac{u(D,\hat{D})}{\sigma}\right)} \exp\left(\frac{u(D,\hat{D}) - u(D',\hat{D})}{\sigma}\right) \\
&\leqslant \frac{\sum_{\hat{D}} \exp\left(\frac{\Delta u}{\sigma}\right) \exp\left(\frac{u(D,\hat{D})}{\sigma}\right)}{\sum_{\hat{D}} \exp\left(\frac{u(D,\hat{D})}{\sigma}\right)} \exp\left(\frac{\Delta u}{\sigma}\right) \\
&\leqslant \exp\left(\frac{2\Delta u}{\sigma}\right) \leqslant \mathrm{e}^{\varepsilon}
\end{aligned}$$

至于 (α, β)-精确, 事实上在 $\mathcal{A}$ 的返回值不是实数后我们还没有定义 (α, β)-精确到底是什么, 所以在这里重新定义一下: 称 $\mathcal{A}$ 满足 (α, β)-精确, 如果

$$\mathbb{P}\left[u(D, \mathcal{A}(D)) \leqslant u^* - \alpha\right] \leqslant \beta$$

其中 $u^* = \max_{D,\hat{D}} u(D,\hat{D})$. 在 BLR Mechanism 中有 $u^* = 0$.

考虑 $\mathcal{A}$ 的像空间 $\mathcal{X}^m$ 中的一个元素 $\hat{D}$, 称 $\hat{D} \in G$ 如果 $u(D,\hat{D}) \geqslant u^* - \frac{\alpha}{2}$, 称 $\hat{D} \in B$ 如果 $u(D,\hat{D}) \leqslant u^* - \alpha$. G 和 B 分别表示 good 和 bad.

引理 10.12. 如果能证明 $G \neq \varnothing$, 则

证明.

$$\begin{aligned}\mathbb{P}\left[u(D,\mathcal{A}(D)) \leqslant u^* - \alpha\right] = \mathbb{P}\left[\mathcal{A}(D) \in B\right] \leqslant \frac{\mathbb{P}\left[\mathcal{A}(D) \in B\right]}{\mathbb{P}\left[\mathcal{A}(D) \in G\right]} &\leqslant \frac{\exp\left(\frac{u^*-\alpha}{\sigma}\right) \cdot |B|}{\exp\left(\frac{u^*-\frac{\alpha}{2}}{\sigma}\right) \cdot |G|} \\ &\leqslant \exp\left(-\frac{\alpha}{2\sigma}\right) |\mathcal{X}|^m = \exp\left(-\frac{\alpha n \varepsilon}{4}\right) |\mathcal{X}|^m \leqslant \beta\end{aligned}$$

□

剩下的开摆了. □

11 Reinforcement Learning

定义一些记号

- $\mathcal{S}$: state space
- $\mathcal{A}$: actione space
- $S_t \in \mathcal{S}, A_t \in \mathcal{A}, R_t \in \mathbb{R}$: state, action, and rewrad at time t
- $P_{s,s'}^a = \mathbb{P}(S_{t+1} = s' | S_t = s, A_t = a)$: transition probability
- $R(s,a) = \mathbb{E}[R_{t+1} | S_t = s, A_t = a]$: reward function
- $\gamma \in (0,1)$: discount factor
- $\pi : \mathcal{S} \to \Delta(\mathcal{A})$: policy, where $\Delta(\mathcal{A})$ denotes the space of probability distribution over $\mathcal{A}$
- $v^\pi(s) = \mathbb{E}\left[\sum_{k \geqslant 0} \gamma^k R_{t+k} \middle| S_t = s\right]$: (state) value function, where probability is over **(i)** the policy π, **(2)** the transition probability $P_{s,s'}^a$.

命题 11.1 (**Bellman Expectation Equation**).

$$v^\pi(s) = \mathbb{E}_{a\sim\pi(s)}\left[R(s,a) + \gamma \sum_{s' \in \mathcal{S}} P_{s,s'}^a v^\pi(s')\right]$$

考虑在固定 policy π 下的 Bellman Expectation Operator $\Phi : \mathbb{R}^N \to \mathbb{R}^N$, 满足

$$\Phi(\mathbf{v})(s) = \mathbb{E}_{a\sim\pi(s)}\left[R(s,a) + \gamma \sum_{s' \in \mathcal{S}} P_{s,s'}^a \mathbf{v}(s')\right]$$

则可以证明 Φ 是一个无穷范数下的 γ-压缩映射 (contraction mapping).

这说明随便找一个 $\mathbf{v}_0 \in \mathbb{R}^N$, 不断对它做 Φ 这个 operator, 它都会收敛到唯一的不动点, 这个点就是 π 的 value function v^π.

11.1 Finding Optimal Policy

定义 11.2 (**Bellman Operator**). 无关 policy π, 定义 Bellman Operator $\Phi^* : \mathbb{R}^N \to \mathbb{R}^N$, 满足

$$\Phi(\mathbf{v})(s) = \max_{a\in\mathcal{A}}\left[R(s,a) + \gamma \sum_{s' \in \mathcal{S}} P_{s,s'}^a \mathbf{v}(s')\right]$$

定理 11.3. Bellman Operator Φ^* 是无穷范数下的 γ-压缩映射.

证明. 考虑 $\mathbf{u}, \mathbf{v} \in \mathbb{R}^N$, 记

$$a_{\mathbf{u}} = \arg\max_a \left[R(s,a) + \gamma \sum_{s' \in \mathcal{S}} P_{s,s'}^a \mathbf{u}(s')\right]$$

$$a_{\mathbf{v}} = \arg\max_a \left[R(s,a) + \gamma \sum_{s' \in \mathcal{S}} P_{s,s'}^a \mathbf{v}(s')\right]$$

此时有 (不妨设 $\Phi^*(\mathbf{v})(s) \geqslant \Phi^*(\mathbf{u})(s)$)

$$\begin{aligned}
\Phi^*(\mathbf{v})(s) - \Phi^*(\mathbf{u})(s) &= \left[R(s, a_{\mathbf{v}}) + \gamma \sum_{s' \in \mathcal{S}} P_{s,s'}^{a_{\mathbf{v}}} \mathbf{v}(s')\right] - \left[R(s, a_{\mathbf{u}}) + \gamma \sum_{s' \in \mathcal{S}} P_{s,s'}^{a_{\mathbf{u}}} \mathbf{u}(s')\right] \\
&\leqslant \left[R(s, a_{\mathbf{v}}) + \gamma \sum_{s' \in \mathcal{S}} P_{s,s'}^{a_{\mathbf{v}}} \mathbf{v}(s')\right] - \left[R(s, a_{\mathbf{v}}) + \gamma \sum_{s' \in \mathcal{S}} P_{s,s'}^{a_{\mathbf{v}}} \mathbf{u}(s')\right] \\
&= \gamma \sum_{s' \in \mathcal{S}} P_{s,s'}^{a_{\mathbf{v}}} (\mathbf{v}(s') - \mathbf{u}(s')) \\
&\leqslant \gamma \sum_{s' \in \mathcal{S}} P_{s,s'}^{a_{\mathbf{v}}} \|\mathbf{v} - \mathbf{u}\|_\infty \\
&= \gamma \|\mathbf{v} - \mathbf{u}\|_\infty
\end{aligned}$$

从而也有 $\|\Phi^*(\mathbf{v}) - \Phi^*(\mathbf{u})\|_\infty \leqslant \gamma \|\mathbf{v} - \mathbf{u}\|_\infty$ □

定理 11.4. 对于任意的 policy π_0, 记 v^{π_0} 为其 value function, 则 $\Phi^*(v^{\pi_0})(s) \geqslant v^{\pi_0}(s)$ 对任意 $s \in \mathcal{S}$ 成立.

证明.

$$\begin{aligned}
v^{\pi_0}(s) &= \mathbb{E}_{a \sim \pi_0(s)} \left[R(s, a) + \gamma \sum_{s' \in \mathcal{S}} P_{s,s'}^{a} v^{\pi_0}(s')\right] \\
\Phi^*(v^{\pi_0})(s) &= \max_{a \in \mathcal{A}} \left[R(s, a) + \gamma \sum_{s' \in \mathcal{S}} P_{s,s'}^{a} v^{\pi_0}(s')\right]
\end{aligned}$$

一目了然了属于是. □

现在有一个问题: 我们有一个 policy π_0, 我们可以对 π_0 的 value function v^{π_0} 作用 Φ^* 得到 $\Phi^*(v^{\pi_0})$, 但是 $\Phi^*(v^{\pi_0})$ 可能不是任何一个 policy 的 value function.

可以从两个层面入手: 首先, 考虑 Φ^* 的唯一不动点 $\mathbf{v}^*$, 它必然是某个 optimal policy π^* 的 value function, 因为只要取 $\pi^*(s) = \arg\max_{a \in \mathcal{A}} \left[R(s, a) + \gamma \sum_{s' \in \mathcal{S}} P_{s,s'}^{a} \mathbf{v}^*(s')\right]$ 即可.

其次, 对于任意的 $\mathbf{v} \in \mathbb{R}^N$, 我们都可以定义关于 $\mathbf{v}$ 的 greedy policy π, $\pi(s) = \arg\max_{a \in \mathcal{A}} \left[R(s, a) + \gamma \sum_{s' \in \mathcal{S}} P_{s,s'}^{a} \mathbf{v}(s')\right]$.

现在, 我们从初始 policy π_0 出发, 记 $\mathbf{v}_0 = v^{\pi_0}$ 为 π_0 的 value function, 迭代计算 $\mathbf{v}_{k+1} = \Phi^*(\mathbf{v}_k)$, 再记 π_k 为 $\mathbf{v}_k$ 诱导的 greedy policy, v^{π_k} 为 π_k 的 value function.

定理 11.5. $\|v^{\pi_k} - \mathbf{v}^*\|_\infty \leqslant \frac{\gamma}{1-\gamma} \|\mathbf{v}_k - \mathbf{v}^*\|_\infty$.