第五讲 Rademacher复杂度、VC维和稳定性

牟克典

2021年4月14日

概要

1. Rademacher复杂度

- 给定假设$h \in H$和样本$S = \{(x_i, y_i)\}_{i=1}^m$, $y_i \in \{+1, -1\}$, 则h的经验误差或者风险

$$
\begin{aligned}
\hat{R}(h) &= \frac{1}{m}\sum_{i=1}^{m}[I(h(x_i) \neq y_i)]. \\
&= \frac{1}{m}\sum_{i=1}^{m}\frac{1 - y_i h(x_i)}{2} \\
&= \frac{1}{2} - \frac{1}{2m}\sum_{i=1}^{m} y_i h(x_i)
\end{aligned}
$$

- $\underset{h \in H}{\operatorname{argmin}}\, \hat{R}(h) = \underset{h \in H}{\operatorname{argmax}}\, \frac{1}{m}\sum_{i=1}^{m} y_i h(x_i)$
- 现实学习任务中要考虑噪声对标记的影响:
  - y_i未必是x_i的真实标记.

- 引入相互独立的Raemacher随机变量σ_i: σ_i等概率取值$+1$, -1, 定义
$$E_\sigma[\sup_{h\in H}\frac{1}{m}\sum_{i=1}^{m}\sigma_i h(x_i)]$$
来体现假设空间的表达能力(对随机噪声的拟合能力).
- 考虑实值函数空间: $\mathcal{F}:\mathcal{Z}\to R$, 令 $Z=\{z_i\in\mathcal{Z}\}_{i=1}^{m}$. 则函数空间 $\mathcal{F}$关于Z的经验Rademacher复杂度
$$\hat{\mathcal{R}}_Z(\mathcal{F})=E_\sigma[\sup_{f\in\mathcal{F}}\frac{1}{m}\sum_{i=1}^{m}\sigma_i f(z_i)]$$
- 经验Rademacher复杂度衡量了 函数空间$\mathcal{F}$与随机噪声在集合Z中的相关性。

- 函数空间 $\mathcal{F}$关于Z的Rademacher复杂度定义为：

$$\mathcal{R}_m(\mathcal{F}) = E_{Z\sim D^m}[\hat{\mathcal{R}}_Z(\mathcal{F})],$$

其中D为$\mathcal{Z}$上的分布.

定理1.1

对实值函数空间: $\mathcal{F}:\mathcal{Z}\to[0,1]$, 依分布$D$从$\mathcal{Z}$中独立同分布采样得到的$Z=\{z_i\}_{i=1}^m$, $0<\delta<1$, 对任意$f\in\mathcal{F}$, 以至少$1-\delta$的概率有

$$E[f(z)] \le \frac{1}{m}\sum_{i=1}^m f(z_i) + 2\mathcal{R}_m(\mathcal{F}) + \sqrt{\frac{\log(\frac{1}{\delta})}{2m}},$$

$$E[f(z)] \le \frac{1}{m}\sum_{i=1}^m f(z_i) + 2\hat{\mathcal{R}}_Z(\mathcal{F}) + 3\sqrt{\frac{\log(\frac{2}{\delta})}{2m}}.$$

引理1.2

对假设空间$H : \mathcal{X} \to \mathcal{Y} = \{+1, -1\}$, 令

$$G = \{(x, y) \mapsto I(h(x) \neq y) : h \in H\},$$

且对样本集$S = \{(x_i, y_i)\}_{i=1}^m$, 令$S_\mathcal{X} = \{x_i\}_{i=1}^m$, 则

$$\hat{\mathcal{R}}_S(G) = \frac{1}{2}\hat{\mathcal{R}}_{S_\mathcal{X}}(H).$$

证明：

$$\begin{aligned}\hat{\mathcal{R}}_S(G) &= E_\sigma[\sup_{h\in H} \frac{1}{m}\sum_{i=1}^m \sigma_i I(h(x_i) \neq y_i)] \\ &= E_\sigma[\sup_{h\in H} \frac{1}{m}\sum_{i=1}^m \sigma_i \frac{1 - y_i h(x_i)}{2}]\end{aligned}$$

证明：

$$
\begin{aligned}
\hat{\mathcal{R}}_S(G) &= E_\sigma[\sup_{h\in H}\frac{1}{m}\sum_{i=1}^{m}\sigma_i I(h(x_i)\neq y_i)] \\
&= E_\sigma[\sup_{h\in H}\frac{1}{m}\sum_{i=1}^{m}\sigma_i\frac{1-y_i h(x_i)}{2}] \\
&= \frac{1}{2}E_\sigma[\sup_{h\in H}\frac{1}{m}\sum_{i=1}^{m}-\sigma_i y_i h(x_i)] \\
&= \frac{1}{2}E_\sigma[\sup_{h\in H}\frac{1}{m}\sum_{i=1}^{m}\sigma_i h(x_i)] \\
&= \frac{1}{2}\hat{\mathcal{R}}_{S_{\mathcal{X}}}(H).
\end{aligned}
$$

定理1.3

设假设空间$H : \mathcal{X} \to \{+1, -1\}$，依分布$D$从$\mathcal{X}$中独立同分布采样得到的$S = \{x_i\}_{i=1}^m$, $0 < \delta < 1$, 对任意$h \in H$, 以至少$1 - \delta$的概率有

$$R(h) \leq \hat{R}(h) + \mathcal{R}_m(H) + \sqrt{\frac{\log(\frac{1}{\delta})}{2m}},$$

$$R(h) \leq \hat{R}(h) + \hat{\mathcal{R}}_S(H) + 3\sqrt{\frac{\log(\frac{2}{\delta})}{2m}}.$$

证明：可由定理1.1和引理1.2直接得到.

注意到

$$\begin{aligned}\hat{\mathcal{R}}_S(H) &= E_\sigma[\sup_{h\in H}\frac{1}{m}\sum_{i=1}^{m}\sigma_i h(x_i)] \\ &= E_\sigma[\sup_{h\in H}\frac{1}{m}\sum_{i=1}^{m}-\sigma_i h(x_i)] \\ &= -E_\sigma[\inf_{h\in H}\frac{1}{m}\sum_{i=1}^{m}\sigma_i h(x_i)]\end{aligned}$$

对于固定的σ来说，计算

$$\inf_{h\in H}\frac{1}{m}\sum_{i=1}^{m}\sigma_i h(x_i)$$

等价于经验风险最小化策略。这意味着$\hat{\mathcal{R}}_S(H)$不易计算。

概要

1. Rademacher复杂度
2. 增长函数

增长函数(Growth function)

假设集H的增长函数$\Pi_H : \mathrm{N} \to \mathrm{N}$定义如下：

$$\forall m \in \mathrm{N} : \Pi_H(m) = \max_{\{x_i\}_{i=1}^m \subseteq \mathcal{X}} \left| \{(h(x_1), \cdots, h(x_m)) : h \in H\} \right|$$

- 增长函数$\Pi_H(m)$给出了H中假设对m个样本点进行分类的最大可能结果数，刻画了假设空间H的表示能力.
- 增长函数$\Pi_H(m)$不依赖输入空间上的分布D.

定理1.4(Massart引理)

设$A \subseteq \mathrm{R}^m$为有限集，且$r = \max_{x\in A} \| x \|_2$，则

$$E_\sigma[\frac{1}{m} \sup_{x\in A} \sum_{i=1}^{m} \sigma_i x_i] \le \frac{r\sqrt{2\log|A|}}{m}$$

其中σ_i是相互独立的Rademacher随机变量，且$x = (x_1, x_2, \cdots, x_m)$.

证明：对任意$t > 0$，应用Jensen不等式， 可以得到

$$\begin{aligned}\exp(tE_\sigma[\sup_{x\in A} \sum_{i=1}^{m} \sigma_i x_i]) &\le E_\sigma[\exp(t \sup_{x\in A} \sum_{i=1}^{m} \sigma_i x_i)]\\ &= E_\sigma[\sup_{x\in A} \exp(t \sum_{i=1}^{m} \sigma_i x_i)]\end{aligned}$$

证明：对任意$t>0$, 应用Jensen不等式， 可以得到

$$\begin{aligned}
\exp(tE_\sigma[\sup_{x\in A}\sum_{i=1}^m \sigma_i x_i]) &\leq E_\sigma[\exp(t\sup_{x\in A}\sum_{i=1}^m \sigma_i x_i)] \\
&= E_\sigma[\sup_{x\in A}\exp(t\sum_{i=1}^m \sigma_i x_i)] \\
&\leq \sum_{x\in A} E_\sigma[\exp(t\sum_{i=1}^m \sigma_i x_i)] \\
&= \sum_{x\in A} E_\sigma[\prod_{i=1}^m \exp(t\sigma_i x_i)] \\
&= \sum_{x\in A}\prod_{i=1}^m E_{\sigma_i}[\exp(t\sigma_i x_i)]
\end{aligned}$$

Hoeffding引理:

如果随机变量X满足$E[X]=0$, $a\le X\le b$ 且$b>a$, 则对任意$t>0$,

$$E[\exp(tX)]\le\exp(\frac{t^2(b-a)^2}{8}).$$

由Hoeffding引理可得到

$$\begin{aligned}\exp(tE_\sigma[\sup_{x\in A}\sum_{i=1}^m\sigma_i x_i]) &\le \sum_{x\in A}\prod_{i=1}^m E_{\sigma_i}[\exp(t\sigma_i x_i)]\\ &\le \sum_{x\in A}\prod_{i=1}^m \exp(\frac{t^2(2x_i)^2}{8})\\ &= \sum_{x\in A}\exp(\frac{t^2}{2}\sum_{i=1}^m x_i^2)\le\sum_{x\in A}\exp(\frac{t^2r^2}{2})\end{aligned}$$

$$\begin{aligned}\exp(tE_\sigma[\sup_{x\in A}\sum_{i=1}^m \sigma_i x_i]) &\leq \sum_{x\in A}\exp(\frac{t^2r^2}{2}) \\ &= |A|\exp(\frac{t^2r^2}{2}).\end{aligned}$$

两边取对数并除以t得到

$$E_\sigma[\sup_{x\in A}\sum_{i=1}^m \sigma_i x_i] \leq \frac{\log|A|}{t} + \frac{tr^2}{2}.$$

选择$t = \frac{\sqrt{2\log|A|}}{r}$， 则

$$E_\sigma[\sup_{x\in A}\sum_{i=1}^{m}\sigma_i x_i] \le r\sqrt{2\log|A|}.$$

两边除以m得到

$$E_\sigma[\frac{1}{m}\sup_{x\in A}\sum_{i=1}^{m}\sigma_i x_i] \le \frac{r\sqrt{2\log|A|}}{m}. \qquad \square$$

推论1.5

设假设空间$H : \mathcal{X} \to \{+1, -1\}$，则

$$\mathcal{R}_m(H) \leq \sqrt{\frac{2 \log \Pi_H(m)}{m}}.$$

证明： 固定样本集$S = \{x_1, x_2, \cdots, x_m\}$, 我们定义

$$H_{|S} = \{h_{|S} = (h(x_1), \cdots, h(x_m)) | h \in H\},$$

则

$$|H_{|S}| \leq \Pi_H(m)$$

且

$$\| h_{|S} \|_2 \leq \sqrt{m}.$$

依据Rademacher复杂性的定义有

$$\begin{aligned}\mathcal{R}_m(H) &= E_S\left[E_\sigma\left[\sup_{h\in H}\frac{1}{m}\sum_{i=1}^{m}\sigma_i h(x_i)\right]\right]\\ &= E_S\left[E_\sigma\left[\sup_{h_{|S}\in H_{|S}}\frac{1}{m}\sum_{i=1}^{m}\sigma_i h(x_i)\right]\right].\end{aligned}$$

由Massart引理可得

$$\begin{aligned}&E_\sigma\left[\sup_{h_{|S}\in H_{|S}}\frac{1}{m}\sum_{i=1}^{m}\sigma_i h(x_i)\right]\\ \leq\;& \frac{\sqrt{m}\sqrt{2\log|H_{|S}|}}{m}\\ \leq\;& \frac{\sqrt{m}\sqrt{2\log\Pi_H(m)}}{m}\end{aligned}$$

因此

$$
\begin{aligned}
\mathcal{R}_m(H) &= E_S\left[E_\sigma\left[\sup_{h\in H}\frac{1}{m}\sum_{i=1}^m \sigma_i h(x_i)\right]\right] \\
&= E_S\left[E_\sigma\left[\sup_{h_{|S}\in H_{|S}}\frac{1}{m}\sum_{i=1}^m \sigma_i h(x_i)\right]\right] \\
&\leq E_S\left[\frac{\sqrt{m}\sqrt{2\log \Pi_H(m)}}{m}\right] \\
&= \sqrt{\frac{2\log \Pi_H(m)}{m}}. \quad \square
\end{aligned}
$$

推论1.6

设假设空间$H:\mathcal{X}\rightarrow\{+1,-1\}$，对任意$0<\delta<1$，以至少 $1-\delta$ 的概率 对任意$h\in H$有

$$R(h)\leq\hat{R}(h)+\sqrt{\frac{2\log\Pi_H(m)}{m}}+\sqrt{\frac{\log(\frac{1}{\delta})}{2m}}.$$

证明：可由定理1.3和推论1.5直接得到. □

事实上，不借助Rademacher复杂性，也可以直接得到下面以增长函数给出的泛化误差界：

定理1.7

对假设空间$H : \mathcal{X} \to \{+1, -1\}$，$m \in \mathrm{N}, 0 < \epsilon < 1$ 和任意$h \in H$有

$$Pr\left[|R(h) - \hat{R}(h)| > \epsilon \right] \leq 4\Pi_H(2m) \exp(-\frac{m\epsilon^2}{8}).$$

Q: 二者相差多少？

概要

- 对分：二分类问题假设空间H对S中样本进行标记的每种可能结果称为对S的一种对分.
- 打散：若H能实现S上的所有对分，即 $\Pi_H(m) = 2^m$ ，则称S能被假设空间H打散.

VC维（Vapkin-Chervonenkis dimension）

假设空间H的VC维是能被H打散的最大样本集的大小，即

$$VC(H) = \max\{m : \Pi_H(m) = 2^m\}.$$

- $VC(H) = d$是指存在大小为d的S能被H打散，但并意味着所有大小为d(或比d小)的样本集都能被H打散。
- 若存在大小为d样本集可以被H打散, 但不存在能被H打散的大小为$d+1$的样本集, 则H的VC维就是d.

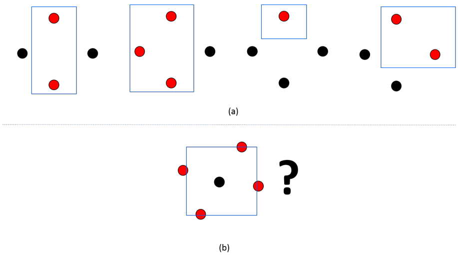

Figure: 轴对齐长方形的VC维

我们考虑由轴对齐长方形构成的假设空间H的VC维

- H可以打散排成菱形的四个点的样本集， 因此$VC(H) \geq 4$.
- H不能打散任何有五个点的样本集， 因此$VC(H) = 4$.

增长函数与VC维的关系：

引理1.8(Sauer引理)

若假设空间H的VC维为d，则对任意$m \in \mathbb{N}$有

$$\Pi_H(m) \leq \sum_{i=0}^{d} \binom{m}{i}.$$

推论1.8

若假设空间H的VC维为d，则对任意$m \geq d$有

$$\Pi_H(m) \leq \left(\frac{em}{d}\right)^d = O(m^d).$$

证明：由Sauer引理可知

$$
\begin{aligned}
\Pi_H(m) &\leq \sum_{i=0}^{d} \binom{m}{i} \\
&\leq \sum_{i=0}^{d} \binom{m}{i} \left(\frac{m}{d}\right)^{d-i} \\
&\leq \sum_{i=0}^{m} \binom{m}{i} \left(\frac{m}{d}\right)^{d-i} \\
&= \left(\frac{m}{d}\right)^{d} \sum_{i=0}^{m} \binom{m}{i} \left(\frac{d}{m}\right)^{i} \\
&= \left(\frac{m}{d}\right)^{d} \left(1+\frac{d}{m}\right)^{m}
\end{aligned}
$$

应用$(1-x)\leq e^{-x}$，可得到

$$\left(1+\frac{d}{m}\right)^m \leq e^{\frac{d}{m}\times m} = e^d.$$

因此

$$\begin{aligned}\Pi_H(m) &\leq \left(\frac{m}{d}\right)^d\left(1+\frac{d}{m}\right)^m \\ &\leq \left(\frac{m}{d}\right)^d e^d \\ &= \left(\frac{em}{d}\right)^d. \quad \square\end{aligned}$$

推论1.9

设假设空间$H:\mathcal{X}\to\{+1,-1\}$，对任意$0<\delta<1$，以至少 $1-\delta$ 的概率 对任意$h\in H$有

$$R(h)\leq\hat{R}(h)+\sqrt{\frac{2d\log\frac{em}{d}}{m}}+\sqrt{\frac{\log(\frac{1}{\delta})}{2m}}.$$

证明：可由推论1.6和推论1.8直接得到. □

由此我们可以得到

$$R(h)\leq\hat{R}(h)+O\left(\sqrt{\frac{\log(m/d)}{m/d}}\right).$$

回顾不借助Rademacher复杂性的泛化误差界：

定理1.7

对假设空间$H: \mathcal{X} \to \{+1, -1\}$， $m \in \mathrm{N}, 0 < \epsilon < 1$ 和任意$h \in H$有

$$Pr\left[|R(h) - \hat{R}(h)| > \epsilon\right] \leq 4\Pi_H(2m)\exp(-\frac{m\epsilon^2}{8}).$$

注意到

$$4\Pi_H(2m)\exp(-\frac{m\epsilon^2}{8}) \leq 4(\frac{2em}{d})^d \exp(-\frac{m\epsilon^2}{8}).$$

令 $4(\frac{2em}{d})^d \exp(-\frac{m\epsilon^2}{8}) = \delta$， 则得到

$$\epsilon = \sqrt{\frac{8d \log \frac{2em}{d} + 8 \log \frac{4}{\delta}}{m}}$$

因此

$$Pr\left[R(h) \leq \hat{R}(h) + \sqrt{\frac{8d \log \frac{2em}{d} + 8 \log \frac{4}{\delta}}{m}}\right] \geq 1 - \delta.$$

概要

1. Rademacher复杂度
2. 增长函数
3. VC维
4. 一致(均匀)稳定性

回顾我们前面关于泛化误差的界

- 对有限假设集H的一致情形来说，

$$R(h_S) \leq \frac{1}{m}(\log|H| + \log\frac{1}{\delta}).$$

- 对有限假设集H来说，

$$\forall h \in H,\ R(h) \leq \hat{R}(h) + \sqrt{\frac{\log|H| + \log\frac{2}{\delta}}{2m}}.$$

- 基于Rademacher复杂度

$$R(h) \leq \hat{R}(h) + \mathcal{R}_m(H) + \sqrt{\frac{\log(\frac{1}{\delta})}{2m}},$$

$$R(h) \leq \hat{R}(h) + \hat{\mathcal{R}}_S(H) + 3\sqrt{\frac{\log(\frac{2}{\delta})}{2m}}.$$

- 基于增长函数

$$R(h) \le \hat{R}(h) + \sqrt{\frac{2\log \Pi_H(m)}{m}} + \sqrt{\frac{\log(\frac{1}{\delta})}{2m}}.$$

- 基于VC维

$$R(h) \le \hat{R}(h) + \sqrt{\frac{2d\log \frac{em}{d}}{m}} + \sqrt{\frac{\log(\frac{1}{\delta})}{2m}}.$$

这些结果与具体学习算法无关，可适用于所有学习算法。

问题：

如何得到与具体学习算法有关的结果？

- 输入空间$\mathcal{X}$， 输出空间$\mathcal{Y} = \{+1, -1\}$
- 我们用z表示标记样例$(x, y) \in \mathcal{X} \times \mathcal{Y}$.
- 训练样本集$S = \{z_i\}_{i=1}^{m}$， 其中每个z_i是来自分布D的独立同分布样本.
- 假设h在点z的损失记为$L_z(h) = L(h(x), y) \in \mathrm{R}_+$.
- 损失函数L的上界为M， 是指对任意 $h \in H$ 和 $z \in \mathcal{X} \times \mathcal{Y}$, 都有 $l_z(h) \leq M$.
- 假设h在S上的经验误差或损失

$$\hat{R}(h) = \frac{1}{m} \sum_{i=1}^{m} L_{z_i}(h).$$

- 假设h的泛化误差

$$R(h) = E_{z \sim D}[L_z(h)].$$

一致稳定性或均匀稳定性(Uniform Stability)

设S和S'是只相差一个样本的任意两个训练样本集. 若学习算法A满足

$$\forall z,\ |L_z(h_S) - L_z(h_{S'})| \leq \beta,$$

则称算法A是一致（或均匀）β-稳定的。

- 直观上，一致稳定的算法A基于相似的训练集学得的模型在相同数据点的损失差异不超过β.
- S只相差一个样本的变化：
  - 移除 S中第i个样例得到的样本集.
  - 替换S中第i个样例得到的样本集.

定理13.1

假定损失函数L的上界为$M \geq 0$. 设学习算法A是一致β-稳定的，且S为依分布D独立同分布采样得到的规模为m的训练样本集，则 以至少$1-\delta$的概率有

$$R(h_S) \leq \hat{R}(h_S) + \beta + (2m\beta + M)\sqrt{\frac{\log\frac{1}{\delta}}{2m}}.$$

证明：我们令

$$\Phi(S) = R(h_S) - \hat{R}(h_S)$$

且S' 为以z'_m替换S中第m个样例z_m得到的样本集， 则

$$|\Phi(S') - \Phi(S)| \leq |R(h_{S'}) - R(h_S)| + |\hat{R}(h_{S'}) - \hat{R}(h_S)|.$$

$$
\begin{aligned}
|R(h_S) - R(h_{S'})| &= |E_z[L_z(h_S)] - E_z[L_z(h_{S'})]| \\
&\leq E_z[|L_z(h_S) - L_z(R_{S'})|] \leq \beta
\end{aligned}
$$

$$
\begin{aligned}
&|\hat{R}(h_S) - \hat{R}(h_{S'})| \\
&= \frac{1}{m}\left|\sum_{i=1}^{m-1}(L_{z_i}(h_S) - L_{z_i}(h_{S'})) + L_{z_m}(h_S) - L_{z'_m}(h_{S'})\right| \\
&\leq \frac{1}{m}[\sum_{i=1}^{m-1}|L_{z_i}(h_S) - L_{z_i}(h_{S'})| + |L_{z_m}(h_S) - L_{z'_m}(h_{S'})|] \\
&\leq \frac{(m-1)\beta}{m} + \frac{M}{m} \leq \beta + \frac{M}{m}
\end{aligned}
$$

由此得到 $|\Phi(S') - \Phi(S)| \leq \beta + \beta + \frac{M}{m} = \frac{2m\beta+M}{m}$.

由McDiarmid不等式可以得到

$$Pr[\Phi(S) \leq \epsilon + E_S[\Phi(S)]] \geq 1 - \exp(\frac{-2m\epsilon^2}{(2m\beta + M)^2})$$

令$\delta = \exp(\frac{-2m\epsilon^2}{(2m\beta+M)^2})$, 得到以不小于$1 - \delta$的概率

$$\Phi(S) \leq E_S[\Phi(S)] + (2m\beta + M)\sqrt{\frac{\log\frac{1}{\delta}}{2m}}.$$

即

$$R(h_S) \leq \hat{R}(h_S) + E_S[\Phi(S)] + (2m\beta + M)\sqrt{\frac{\log\frac{1}{\delta}}{2m}}.$$

现在考虑 $E_S[\Phi(S)] = E_S[R(h_S) - \hat{R}(h_S)]$.

$$E_{S\sim D^m}[R(h_S)] = E_{S\sim D^m}[E_{z\sim D}[L_z(h_S)]] = E_{S,z\sim D^{m+1}}[L_z(h_S)]$$

$$E_{S\sim D^m}[\hat{R}(h_S)] = \frac{1}{m}\sum_{i=1}^{m} E_{S\sim D^m}[L_{z_i}(h_S)] = E_{S\sim D^m}[L_{z_1}(h_S)]$$

令S'是从S和z构成的$m+1$点中抽取的含z的m个点的样本集，则

$$E_{S\sim D^m}[\hat{R}(h_S)] = E_{S\sim D^m}[L_{z_1}(h_S)] = E_{S,z\sim D^{m+1}}[L_z(h_{S'})].$$

考虑一致稳定性，可得到

$$\begin{aligned}|E_S[\Phi(S)]| &= |E_{S,z\sim D^{m+1}}[L_z(h_S)] - E_{S,z\sim D^{m+1}}[L_z(h_{S'})]| \\ &\leq E_{S,z\sim D^{m+1}}[|L_z(h_S) - L_z(h_{S'})|] \\ &\leq E_{S,z\sim D^{m+1}}[\beta] = \beta.\end{aligned}$$

注意到

$$E_S[\Phi(S)] \le |E_S[\Phi(S)]| \le \beta,$$

因此

$$R(h_S) \le \hat{R}(h_S) + \beta + (2m\beta + M)\sqrt{\frac{\log \frac{1}{\delta}}{2m}}.\square$$

小结

- Rademacher复杂度
  - Rademacher随机变量
  - 经验Rademacher复杂度
  - Rademacher复杂度
- 增长函数
- VC维
  - 对分
  - 打散
- 稳定性