LECTURE 28

Example (a): half-lives $\mathcal{F} = \{ \mathbb{1}_{(-\infty, a]} : a \in \mathbb{R} \}$. $\quad vc(\mathcal{F}) = 1$.

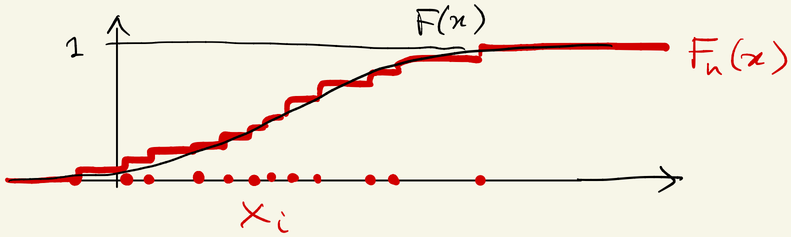
$\forall f \in \mathcal{F}$:

$\mathbb{E} f(X) = \mathbb{P}\{X \le a\} =: F(a)$, the cumulative distribution function (CDF) of X

$\frac{1}{n} \sum_{i}^{n} f(X_i) = \frac{\#\{X_i \le a\}}{n} =: F_n(a)$, the empirical CDF of X.

$\Rightarrow$ ULLN yields $\quad \mathbb{E} \sup_{a \in \mathbb{R}} |F_n(a) - F(a)| \le C\sqrt{\frac{\log n}{n}}$, i.e:

> **Glivenko–Cantelli Thm** $\;\forall$ random variable X, the CDF of X can be uniformly estimated from a sample of n pts:
> $$\mathbb{E} \| F_n - F \|_\infty \le C\sqrt{\frac{\log n}{n}}.$$



All quantiles of X can be estimated from one sample $X_1, \dots, X_n$.

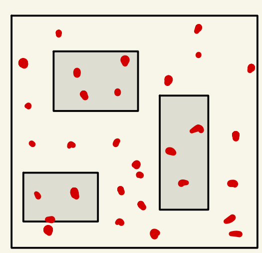
Example (b): axis-aligned rectangles $\mathcal{F} = \{ \mathbb{1}_R : R = [a,b] \times [c,d] \}$

$vc(\mathcal{F}) = 4$

let $X \sim \text{Unif}\,[0,1]^2 \Rightarrow$

$\mathbb{E} f(X) = \text{Area}(R)$; $\quad \frac{1}{n} \sum_{i=1}^{n} f(X) = \frac{\#\{X_i \in R\}}{n} \quad \overset{\text{ULLN}}{\Longrightarrow}$

$$\boxed{\mathbb{E} \sup_R \left| \frac{\#\{X_i \in R\}}{n} - \text{Area}(R) \right| \le C\sqrt{\frac{\log n}{n}}}$$



discrepancy

uniformly over all rectangles

recovering the discrepancy thm (lec. 6)

Example (c): A similar result holds for other common shapes: triangles, circles, ellipses, etc. $\quad$ since $vc(\mathcal{F}) < \infty$.

BACK TO MACHINE LEARNING

- Recall ERM from lec. 24: we want to predict $Y = Y(X)$ from X

Risk (test error): $R(h) := \mathbb{E}\,\ell(h(X), Y)$, $\quad h^* = \operatorname{argmin}_{h \in \mathcal{H}} R(h)$.

Empirical risk (training error): $\underbrace{R_n(h) := \frac{1}{n}\sum_{i=1}^{n} \ell(h(X_i), Y_i), \quad h_n^* := \operatorname{argmin}_{h \in \mathcal{H}} R_n(h)}_{\text{"ERM}}$

- Generalization Bound:

$$R(h_n^*) \le R(h^*) + 2 \sup_{h \in \mathcal{H}} \underbrace{\left| R_n(h) - R(h) \right|}_{\parallel}$$

$$\frac{1}{n}\sum_{i=1}^{n} \underbrace{\ell(h(X_i), Y_i)}_{\parallel} - \mathbb{E}\,\ell(h(X), Y)$$

$f(X_i)$, apply ULLN $\Rightarrow$

THM (VC Generalization bound)

> If $\mathcal{H}$ is a Boolean class with $d = vc(\mathcal{H}) < \infty$, then
> $$\mathbb{E} R(h_n^*) \le R(h^*) + C\sqrt{\frac{d \log n}{n}}$$

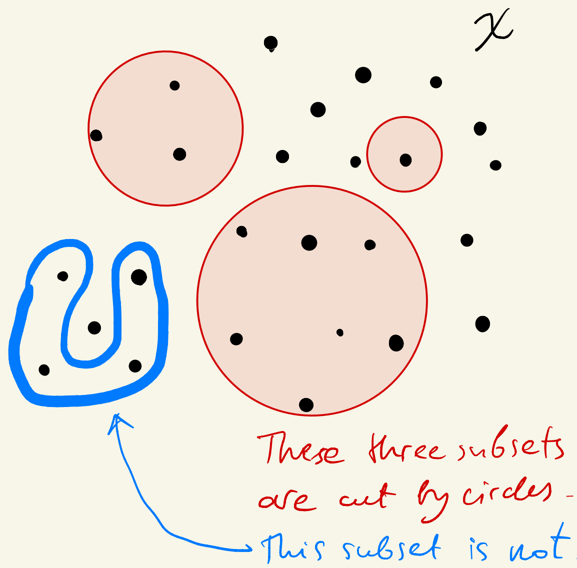
generalization error

Remark: $\log n$ can be removed (see my book)

$\Rightarrow$ generalizes well from a sample of size

$$\boxed{n \sim vc(\mathcal{H})}$$

One More Applications of Sauer-Shelah Lemma



① Consider any n-point set $\mathcal{X} \subset \mathbb{R}^2$.
How many subsets of $\mathcal{X}$ are cut by circles?

$$\boxed{\textbf{Prop} \quad \text{At most } \sum_{k=0}^{3} \binom{n}{k}.} \approx \frac{n^3}{6} \text{ for large } n$$

Proof: Consider the class of functions

$$\mathcal{F} = \{ \mathbb{1}_C : C \text{ is a circle in } \mathbb{R}^2 \}$$

$vc(\mathcal{F}) = 3$.

Restrict the domain of each function $f \in \mathcal{F}$ to $\mathcal{X}$ $\Rightarrow$ $\mathcal{F}|_{\mathcal{X}}$

$=$

$\{ \mathbb{1}_{C \cap \mathcal{X}} : C \text{ is a circle} \}$

indicators of the subsets of $\mathcal{X}$ cut by circles

$vc(\mathcal{F}|_{\mathcal{X}}) \le vc(\mathcal{F}) \le 3$. Sauer-Shelah lemma $\Rightarrow$

$$|\mathcal{F}|_{\mathcal{X}}| \le \sum_{k=0}^{3} \binom{n}{k}.$$

Remark Prop is optimal for most sets $\mathcal{X}$
(whenever no 4 pts lie on the same circle)