Vapnik–Chervonenkis Classes (wrap-up)

Dudley classes:

$\mathcal{G}$ – a class of fcns $g : \mathcal{Z} \to \mathbb{R}$

$h : \mathcal{Z} \to \mathbb{R}$ – arbitrary fcn (not necc. in $\mathcal{G}$)

$$\text{pos}(g+h) := \{z \in \mathcal{Z} : g(z) + h(z) \geq 0\}$$

$$\mathcal{G} \longrightarrow \text{pos}(\mathcal{G}+h) := \{\text{pos}(g+h) : g \in \mathcal{G}\}$$

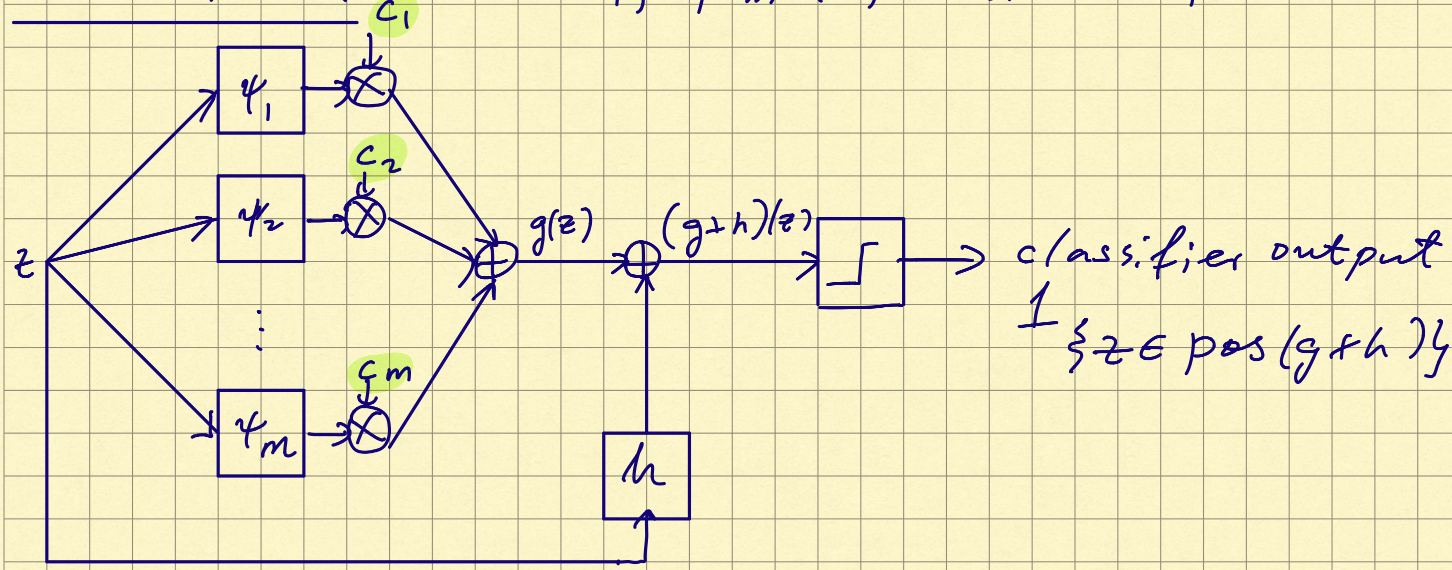
$\mathcal{G}$: an m-dim. linear space

$\exists$ fcns $\psi_1, \dots, \psi_m : \mathcal{Z} \to \mathbb{R}$ s.t. any $g \in \mathcal{G}$ can be uniquely represented as

$$g(z) = \sum_{j=1}^{m} c_j \psi_j(z)$$

for some $c_1, \dots, c_m \in \mathbb{R}$.

Classifiers: $c_1, \dots, c_m$: tunable coefficients



Thm (Dudley) If $\mathcal{G}$ is an m-dim. linear space of functions, then

$$V(\text{pos}(\mathcal{G}+h)) = m, \quad \forall h.$$

Examples

1) linear classifiers $\quad \mathcal{Z} = \mathbb{R}^d$

$$\mathcal{G} := \{ z \mapsto \langle w, z \rangle + b : \ w \in \mathbb{R}^d \setminus \{0\},\ b \in \mathbb{R} \}$$

$$\mathrm{pos}(\mathcal{G}) = \{ \{ z \in \mathbb{R}^d : \langle w, z \rangle + b \geq 0 \} \}$$

$$g(z) = \langle w, z \rangle + b \qquad w = (w_1, \dots, w_d)$$
$$\qquad z = (z_1, \dots, z_d)$$

$$= \sum_{j=1}^{d} w_j \cdot z_j + b$$

$$= \sum_{j=1}^{d+1} c_j \cdot \psi_j(z) \quad \text{where} \quad c_j = \begin{cases} w_j, & j \in [d] \\ b, & j = d+1 \end{cases}$$

$$\psi_j(z) = \begin{cases} z_j, & j \in [d] \\ 1, & j = d+1 \end{cases}$$

$\Rightarrow$ $\mathcal{G}$ is a linear space spanned by $1, z_1, \dots, z_d$

$$V(\mathrm{pos}(\mathcal{G})) = d+1$$

2) indicators of closed balls in $\mathbb{R}^d$: $\quad \mathcal{Z} = \mathbb{R}^d$

$$\tilde{g}(z) = r^2 - \|z - x\|^2 \qquad r \in \mathbb{R}$$
$$\qquad x \in \mathbb{R}^d$$

$\mathrm{pos}(\tilde{g}) := B_r(x)$ – closed ball in $(\mathbb{R}^d, \|\cdot\|)$ of rad. r centered at x

$$V(\text{balls in } \mathbb{R}^d) = d+1$$

$$\tilde{g}(z) = r^2 - \|z - x\|^2$$
$$= r^2 - \sum_{j=1}^{d} (z_j - x_j)^2$$

$$= r^2 - \sum_{j=1}^{d} \left(z_j^2 - 2x_j z_j + x_j^2 \right)$$

$$= \underbrace{\sum_{j=1}^{d} 2x_j z_j + r^2 - \sum_{j=1}^{d} x_j^2}_{g(z)} - \sum_{j=1}^{d} z_j^2$$

$$= \sum_{j=1}^{d+1} c_j \psi_j(z) + h(z)$$

where :

$$c_j = \begin{cases} x_j, & j \in [d] \\ r^2 - \sum_{j=1}^{d} x_j^2 & j = d+1 \end{cases}$$

$$\psi_j(z) = \begin{cases} 2z_j, & j \in [d] \\ 1, & j = d+1 \end{cases}$$

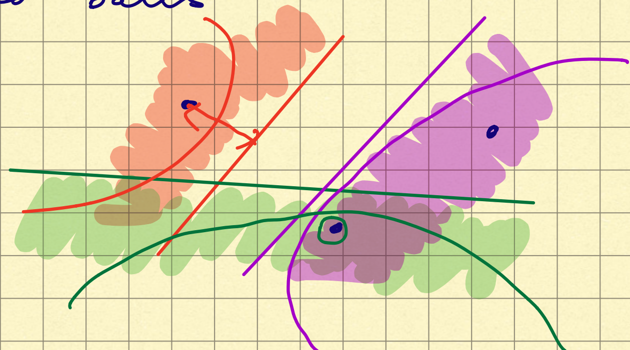
$$h(z) = -\sum_{j=1}^{d} z_j^2$$

$$\mathcal{G} \subseteq \operatorname{span}\{\psi_1, \dots, \psi_{d+1}\} \Rightarrow \dim(\mathcal{G}) \leq d+1$$

[$c_1, \dots, c_{d+1}$ are not free: $c_1 = x_1, \dots, c_d = x_d$,
but $c_{d+1} = r^2 - \sum_{j=1}^{d} x_j^2 = r^2 - \sum_{j=1}^{d} c_j^2$]

By Dudley, $V(\operatorname{pos}(\mathcal{G} + h))$
$\leq V(\operatorname{pos}(\operatorname{span}\{\psi_1, \dots, \psi_{d+1}\} + h)) = d+1.$

Can shatter a set of $d+1$ pts in $\mathbb{R}^d$ using closed balls



shattering by half-spaces $\Rightarrow$ shattering by ball

Growth of shatter coefficients (Sauer-Shelah lemma)

$V(\mathcal{F}) = d < \infty$ ($\mathcal{F}$: $\{0,1\}$-valued fcns on $\mathcal{Z}$)

$$S_n(\mathcal{F}) = \begin{cases} 2^n, & n \le d \\ \text{poly}(n), & n > d \end{cases} \qquad \sim (n+1)^d$$

$$(n+1)^d = o(2^n)$$

Notation:

$$\binom{n}{\le d} := \#\{\text{subsets of card.} \le d \text{ of a set of cardinality } n\}$$

$$= \begin{cases} 2^n & , \ d \ge n \\ \sum_{i=0}^{d} \binom{n}{i} & , \ d < n \end{cases}$$

$$= \sum_{i=0}^{d} \binom{n}{i} \qquad \binom{n}{i} = 0 \text{ if } i > n$$

Sauer-Shelah Lemma Let $\mathcal{F}$ be a class of binary-valued fcns, s.t. $V(\mathcal{F}) = d < \infty$. Then

$$S_n(\mathcal{F}) \le \binom{n}{\le d}.$$

Useful bounds:

1) $\binom{n}{\le d} \le (n+1)^d$ (for all $n, d \ge 1$)

2) $\binom{n}{\le d} \le \left(\frac{ne}{d}\right)^d$ for $n \ge d$

Proofs

1) $\boxed{0 \,|\, 1 \,|\, 2 \,|\, 3 \,|\, \cdots \,|\, n}$

draw d elements of $\{0, 1, \ldots, n\}$ with replacement

draw $\underbrace{(0, 5, 3, 3, 0, 0)}_{6 \text{ draws}} \longrightarrow \{3, 5\}$

$(7, 1, 1, 0, 0, 8) \longrightarrow \{1, 7, 8\}$

2) $n \geq d$ $\qquad d/n \leq 1$

$\left(\frac{d}{n}\right)^d \leq \left(\frac{d}{n}\right)^i \quad \forall i \leq d$

$$\left(\frac{d}{n}\right)^d \binom{n}{\leq d} = \left(\frac{d}{n}\right)^d \sum_{i=0}^{d} \binom{n}{i}$$

$$\leq \sum_{i=0}^{d} \binom{n}{i} \left(\frac{d}{n}\right)^i$$

$$\leq \sum_{i=0}^{n} \binom{n}{i} \left(\frac{d}{n}\right)^i$$

$$= \left(1 + \frac{d}{n}\right)^n \qquad \text{(binomial thm)}$$

$$\leq e^d$$

$\log x \leq x - 1$

$x \leq e^{x-1}$

$$\Rightarrow \mathbb{S}_n(\mathcal{F}) \leq \binom{n}{\leq d} \leq \left(\frac{ne}{d}\right)^d$$

Main Consequence: if $\mathcal{F}$ is a class of $\{0,1\}$-valued fcns on $\mathcal{Z}$, and $V(\mathcal{F}) = d < \infty$, then $\forall z^n \in \mathcal{Z}^n$,

$$R_n(\mathcal{F}(z^n)) \leq 2\sqrt{\frac{\log |\mathbb{S}_n(\mathcal{F})|}{n}} \qquad \text{(Finite Class Lemma)}$$

$$\leq 2\sqrt{\frac{d \log(n+1)}{n}} \qquad \text{(Sauer-Shelah)}$$

$$\text{or } \quad 2\sqrt{\frac{d \log(ne/d)}{n}} \quad \text{for } n \geq d$$

If $V(\mathcal{F}) = d < \infty$, then $R_n(\mathcal{F}(z^n)) = O\left(\sqrt{\frac{d \log n}{n}}\right)$

— in fact, $R_n(\mathcal{F}(z^n)) = O(\sqrt{d/n})$ [Dudley's chaining argument]

Back to Stat. Learning: Binary Classification

$z = (X, Y)$ $\quad X \in \mathcal{X}$ (feature space)

$Y \in \{0,1\}$

• Bayes classifier: $f^*(x) = \mathbb{1}_{\{\eta(x) \geq 1/2\}}$

where $\eta(x) = P[Y=1 \mid X=x]$

$$P(Y \neq f^*(X)) = \min_{f: \mathcal{X} \to \{0,1\}} P(Y = f(X))$$

• learning: $P = \mathcal{L}((X,Y))$ unknown

$Z_1, \dots, Z_n \overset{iid}{\sim} P$

$\mathcal{C}$ — classifiers (subsets of $\mathcal{X}$)

$x \longmapsto \hat{y} = \mathbb{1}_{\{x \in C\}}$

$$\ell_C(z) = \ell_C((x,y)) = \mathbb{1}_{\{y \neq \hat{y}\}} = \mathbb{1}_{\{y \neq \mathbb{1}_{\{x \in C\}}\}}$$

$$\mathcal{C} \longrightarrow \mathcal{F}_{\mathcal{C}} = \{\ell_C : C \in \mathcal{C}\}$$

ERM: $\hat{C}_n = \underset{C \in \mathcal{C}}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} \ell_C((X_i, Y_i))$

Goal: $L(\hat{C}_n) \approx \inf_{C \in \mathcal{C}} L(C)$ w.h.p.

$$L(C) := P\{Y \neq 1_{\{X \in C\}}\}$$

Bounds for ERM: w.p. $\geq 1-\delta$,

$$L(\hat{C}_n) \leq \inf_{C \in \mathcal{C}} L(C) + \underbrace{c}_{\text{abs. const.}} \sqrt{\frac{V(\mathcal{F}_{\mathcal{C}})}{n}} + \sqrt{\frac{2\log(1/\delta)}{n}}$$

[symmetrization, FCL, VC estimates, &c]

Lemma (proof in next lecture)

$$\underbrace{V(\mathcal{F}_{\mathcal{C}})}_{\text{fcns on } \mathcal{X} \times \{0,1\}} = \underbrace{V(\mathcal{C})}_{\text{fcns on } \mathcal{X}}$$