Growth Functions

Shattering Coefficient

$$\mathcal{N}(\mathcal{F}, n) = \max\{|\mathcal{F}_{Z_n}| \mid X_1, ..., X_n \in \mathcal{X}\}.$$

Number of ways that the function space can separate the patterns into two classes

$$\mathcal{N}(\mathcal{F}, n) = 2^n,$$ Z_n is shattered

- Shattering means that there exists a sample of *n* patterns which can be separated in all possible ways

- it does not mean that this applies to all possible samples of *n* patterns.

- Measure of capacity of function class

Generalization Bound

$$R(f) \leq R_{\mathrm{emp}}(f) + \sqrt{\frac{4}{n}\big(\log(2\mathcal{N}(\mathcal{F}, n)) - \log(\delta)\big)}.$$

VC Dimension

We say that a sample Z_n of size n is *shattered by function class* $\mathcal{F}$ if the function class can realize any labeling on the given sample, that is $|\mathcal{F}_{Z_n}| = 2^n$. The *VC dimension of* $\mathcal{F}$, denoted by $\mathrm{VC}(\mathcal{F})$, is now defined as the largest number n such that there exists a sample of size n which is shattered by $\mathcal{F}$. Formally,

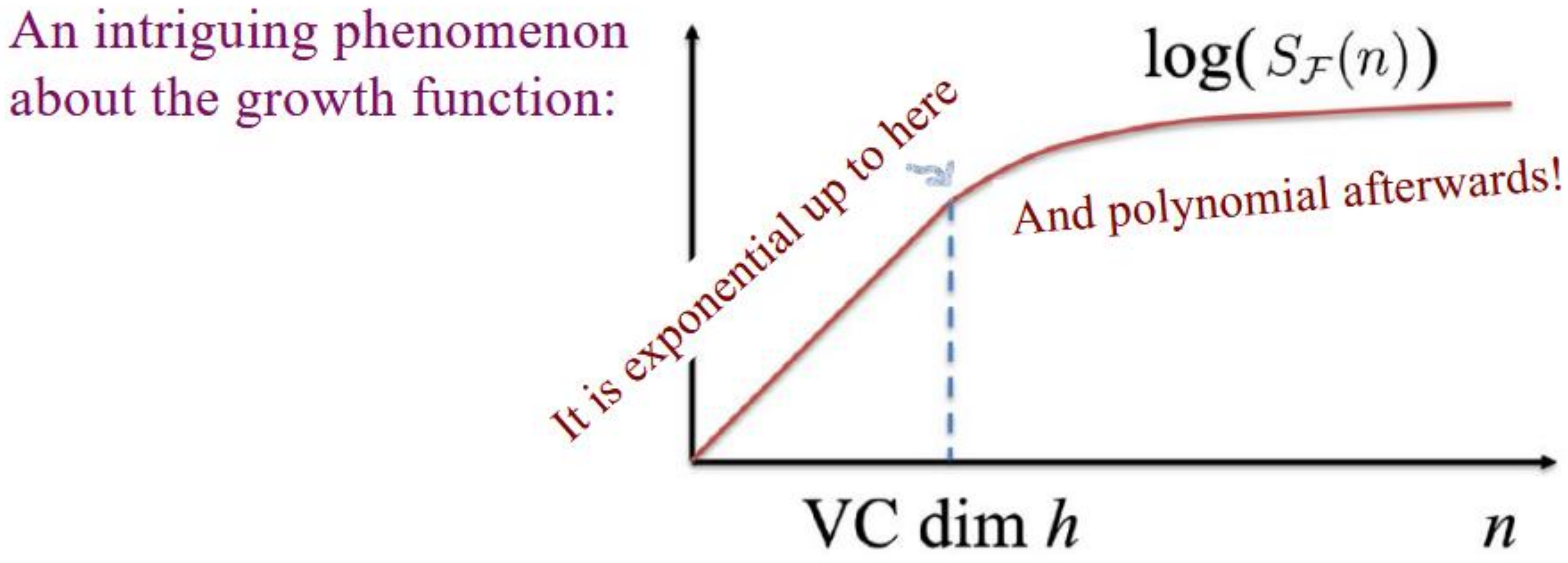
$$\mathrm{VC}(\mathcal{F}) = \max\{n \in \mathbb{N} \mid |\mathcal{F}_{Z_n}| = 2^n \text{ for some } Z_n\}.$$

Growth Function of Shattering Coefficient

Definition: The *growth function* of function class $\mathcal{F}$ is the maximum number of ways into which n points can be classified by the function class.

$$S_{\mathcal{F}}(n) = \sup_{(z_1,\dots,z_n)} |\mathcal{F}_{z_1,\dots,z_n}|$$

Growth Function



Sauer-Shellah Lemma

Lemma 5 (Vapnik, Chervonenkis, Sauer, Shelah) *Let $\mathcal{F}$ be a function class with finite VC dimension d. Then*

$$\mathcal{N}(\mathcal{F}, n) \leq \sum_{i=0}^{d} \binom{n}{i}$$

for all $n \in \mathbb{N}$. In particular, for all $n \geq d$ we have

$$\mathcal{N}(\mathcal{F}, n) \leq \left(\frac{en}{d}\right)^d .$$

VC Bound

Theorem VC-Bound. If $\mathcal{F}$ has VC dim h, and for $n \geq h$, with probability at least $1 - \delta$,

$$\forall f \in \mathcal{F} \quad R^{\text{true}}(f) \leq R^{\text{emp}}(f) + 2\sqrt{2\frac{h(\ln(2n)+1) + \log\frac{4}{\delta}}{n}}$$