Rule of thumb: If $n \gg \mathrm{VCdim}(\mathcal{F})$ then there is a guarantee that $\hat{R}(h)$ is close to $R(h)$ with high probability.

As in the case of threshold functions, the VCdim of $\mathcal{F}$ is often related to the number of parameters required to specify an element in $\mathcal{F}$.
For instance:

Thm.: Let $\mathcal{G} \subseteq \mathbb{R}^{\mathcal{X}}$ be an $\mathbb{R}$-vec. space of functions $f \in \mathbb{R}^{\mathcal{X}}$ and $\mathcal{F} := \{ x \mapsto \mathrm{sgn}[g(x) + f(x)] \mid g \in \mathcal{G} \}$. Then $\mathrm{VCdim}(\mathcal{F}) = \dim(\mathcal{G})$.

proof: ... as in lec. notes.

Corollary: (VC-dim. of set of all half spaces)
$\mathcal{F} := \{ h : \mathbb{R}^d \to \{-1, 1\} \mid \exists (v, b) \in \mathbb{R}^d \times \mathbb{R} : h(x) = \mathrm{sgn}[\langle v, x \rangle - b] \}$
has $\mathrm{VCdim}(\mathcal{F}) = d+1$.

proof: Apply thm. to linear space of functions spanned by
$g_i(x) := x_i$ for $i = 1, \ldots, d$
$g_{d+1}(x) = 1$. $\square$

However, there is no general relation between # parameters & VCdim.
The standard 'counterexample' is $\mathcal{F} := \{ x \mapsto \mathrm{sgn}[\sin(\alpha x)] \mid \alpha \in \mathbb{R} \}$ for which $\mathrm{VCdim}(\mathcal{F}) = \infty$.

Note: the concepts of growth function and VCdim are directly applicable only if $|\mathcal{Y}| < \infty$ and $|\mathcal{Y}| = 2$, respectively. In order to be more general, we need new concepts:

Covering numbers

Def.: (M,d) pseudometric space, $A, B \subseteq M$, $\varepsilon > 0$.

- A is <u>ε-cover</u> of B if $\forall b \in B \; \exists a \in A : d(a,b) \le \varepsilon$.
- <u>ε-covering number</u> $N(\varepsilon, B)$ is the smallest cardinality of any ε-cover of B.
- $A \subseteq B$ is <u>ε-packing</u> of B if $a, b \in A \Rightarrow d(a,b) > \varepsilon$
- <u>ε-packing number</u> $M(\varepsilon, B)$ is the largest cardinality of any ε-packing of B.

Rem.:
- $N(\frac{\varepsilon}{2}, B) \geq M(\varepsilon, B) \geq N(\varepsilon, B)$
- Covering-/packing numbers are often used in information theory / coding theory. One reason is the following:

Prop.: Let B be a subset of a pseudo-metric space and $\beta(\varepsilon, B)$ the smallest number of bits sufficient to specify every $b \in B$ up to an error of at most ε in the pseudo-metric. Then

$$\log_2 N(\varepsilon, B) \le \beta(\varepsilon, B) \le \log_2 M(\varepsilon, B).$$

proof: homework.

Exp.: $\|\cdot\|$-balls in $\mathbb{R}^d$: for $x \in \mathbb{R}^d$ define $B_r(x) := \{ y \in \mathbb{R}^d \mid \|y - x\| \le r \}$ and let $\{x_1, \dots, x_M\} \subseteq \mathbb{R}^d$ be max. ε-packing of $B_r(0)$. So $M = M(\varepsilon, B_r(0))$ and $B_{\frac{\varepsilon}{2}}(x_i)$ are disjoint and lie in $B_{r+\frac{\varepsilon}{2}}(x_i)$. Let $v := \mathrm{vol}(B_1(0))$. Then

$$M(\varepsilon, B_r(0)) \le \frac{\mathrm{vol}(B_{r+\frac{\varepsilon}{2}}(0))}{\mathrm{vol}(B_{\frac{\varepsilon}{2}}(0))} = \frac{(r+\frac{\varepsilon}{2})^d v}{(\frac{\varepsilon}{2})^d v} \overset{\text{if } \varepsilon \le r}{\le} \underline{\underline{\left(\frac{3r}{\varepsilon}\right)^d}}$$

For a bounded object B of algebraic dim. d, $\ln M(\varepsilon, B) \sim d \ln \frac{1}{\varepsilon}$ is the typical scaling.