$1-\delta$, for any $h \in H$,

$$R(h) \leq \widehat{R}(h) + \sqrt{\frac{2(N+1)\log\frac{em}{N+1}}{m}} + \sqrt{\frac{\log\frac{1}{\delta}}{2m}}. \tag{4.37}$$

When the dimension of the feature space N is large compared to the sample size, this bound is uninformative. The following theorem presents instead a bound on the VC-dimension of canonical hyperplanes that does not depend on the dimension of feature space N, but only on the margin and the radius r of the sphere containing the data.

Theorem 4.2
Let $S \subseteq \{\mathbf{x}\colon \|\mathbf{x}\| \leq r\}$. Then, the VC-dimension d of the set of canonical hyperplanes $\{x \mapsto \mathrm{sgn}(\mathbf{w}\cdot\mathbf{x})\colon \min_{x\in S}|\mathbf{w}\cdot\mathbf{x}| = 1 \wedge \|\mathbf{w}\| \leq \Lambda\}$ verifies

$$d \leq r^2\Lambda^2\,.$$

Proof Assume $\{\mathbf{x}_1, \ldots, \mathbf{x}_d\}$ is a set that can be fully shattered. Then, for all $\mathbf{y} = (y_1, \ldots, y_d) \in \{-1, +1\}^d$, there exists $\mathbf{w}$ such that,

$$\forall i \in [1, d], 1 \leq y_i(\mathbf{w}\cdot\mathbf{x}_i)\,.$$

Summing up these inequalities yields

$$d \leq \mathbf{w}\cdot\sum_{i=1}^{d} y_i\mathbf{x}_i \leq \|\mathbf{w}\|\|\sum_{i=1}^{d} y_i\mathbf{x}_i\| \leq \Lambda\|\sum_{i=1}^{d} y_i\mathbf{x}_i\|\,.$$

Since this inequality holds for all $\mathbf{y} \in \{-1, +1\}^d$, it also holds on expectation over $y_1, \ldots, y_d$ drawn i.i.d. according to a uniform distribution over $\{-1, +1\}$. In view of the independence assumption, for $i \neq j$ we have $\mathrm{E}[y_iy_j] = \mathrm{E}[y_i]\,\mathrm{E}[y_j]$. Thus, since the distribution is uniform, $\mathrm{E}[y_iy_j] = 0$ if $i \neq j$, $\mathrm{E}[y_iy_j] = 1$ otherwise. This gives

$$\begin{aligned}
d &\leq \Lambda \mathop{\mathrm{E}}_{\mathbf{y}}[\|\sum_{i=1}^{d} y_i\mathbf{x}_i\|] && \text{(taking expectations)}\\
&\leq \Lambda\Big[\mathop{\mathrm{E}}_{\mathbf{y}}[\|\sum_{i=1}^{d} y_i\mathbf{x}_i\|^2]\Big]^{1/2} && \text{(Jensen's inequality)}\\
&= \Lambda\Big[\sum_{i,j=1}^{d} \mathop{\mathrm{E}}_{\mathbf{y}}[y_iy_j](\mathbf{x}_i\cdot\mathbf{x}_j)\Big]^{1/2}\\
&= \Lambda\Big[\sum_{i=1}^{d}(\mathbf{x}_i\cdot\mathbf{x}_i)\Big]^{1/2} \leq \Lambda\big[dr^2\big]^{1/2} = \Lambda r\sqrt{d}.
\end{aligned}$$

Thus, $\sqrt{d} \leq \Lambda r$, which completes the proof. ■