1 Theorem 2

Theorem 2 (Aussage):

Wir machen die Annahme, dass $||f_0||_\infty \leq F$ für ein $F \geq 1$.
Sei $\hat{f}_n$ ein Schätzer für f_0, der aus $\mathcal{F}(L,p,s,F)$ gewählt wird und sei $\Delta_n(\hat{f}_n, f_0)$ wie oben.
Dann gilt:

$\forall \varepsilon \in (0,1] \; \exists C_\varepsilon > 0 \; \forall n \in \mathbb{N}:$

$$(1-\varepsilon)^2 \Delta_n(\hat{f}_n, f_0) - \tau_{\varepsilon,n} \leq R(\hat{f}_n, f_0) \leq (1+\varepsilon)^2 \left(\inf_{f \in \mathcal{F}(L,p,s,F)} ||f - f_0||_\infty^2 + \Delta_n(\hat{f}_n, f_0) \right) + \tau_{\varepsilon,n}$$

für

$$\tau_{\varepsilon,n} := C_\varepsilon F^2 \frac{(s+1)\log(n(s+1)^L p_0 p_{l+1})}{n}.$$

Zuerst ein paar unbewiesene Hilfslemmata.

Hilfslemma 1

Für einen Schätzer $\hat{f}$ aus der Funktionsfamilie $\mathcal{F} \subset \{f : [0,1]^d \to [0,F]\}$ sei $\Delta_n(f, f_0, \mathcal{F})$ wie oben definiert.
Falls $(\mathcal{F}, ||\cdot||_\infty)$ als metrischer Raum relativkompakt (also das Abschluss sei überdeckungskompakt), dann können wir die Überdeckungszahlen $N(\delta, \mathcal{F}, ||\cdot||_\infty)$ wie folgt definieren.

Für $\delta > 0$ beschreibe $N(\delta, \mathcal{F}, ||\cdot||_\infty)$ die minimale Anzahl an Punkten $x_1, ..., x_{N(\delta,\mathcal{F},||\cdot||_\infty)} \in \mathcal{F}$, sodass $\mathcal{F} \subset \bigcup_{i=1}^{N(\delta,\mathcal{F},||\cdot||_\infty)} B_\delta(x_i)$ gilt.
Nach relativkompaktheit ist $N(\delta, \mathcal{F}, ||\cdot||_\infty)$ immer endlich.

Die Aussage des Hilfslemmas ist nun, dass für jedes $\delta \in (0,1]$ mit $N(\delta, \mathcal{F}, ||\cdot||_\infty) \geq 3$ gilt:

$$\begin{aligned}&(1-\varepsilon)^2 \Delta_n(\hat{f}, f_0) - F^2 \frac{18\log(N(\delta, \mathcal{F}, ||\cdot||_\infty)) + 76}{n\varepsilon} - 38\,\delta F \\ &\leq R(\hat{f}, f_0)\end{aligned}$$

$$\leq (1+\varepsilon)^2 \Bigg(\underbrace{\inf_{f \in \mathcal{F}} E[(f(X) - f_0(X))^2]}_{\text{bestes Risiko}} + F^2 \frac{18 \log(N(\delta, \mathcal{F}, || \cdot ||_\infty)) + 72}{n\varepsilon} + 32\,\delta F + \Delta_n(\hat{f}, f_0) \Bigg)$$

(Für den Beweis siehe Hieber Appendix C Seite 10-14)

Hilfslemma 2

Definiere $V := \prod_{l=0}^{L+1} (p_l + 1)$, dann gilt für jedes $\delta > 0$:

$$\log(N(\delta, \mathcal{F}(L, p, s, \infty), || \cdot ||_\infty)) \leq (s+1) \log \left(2 \frac{L+1}{\delta} V^2 \right).$$

(Für den Beweis siehe Hieber Appendix C Seite 15-16)

Theorem 2 (Beweis):

Eine genauere Betrachtung der Definition von $\mathcal{F}(L, p, s)$ führt zu der Aussage

$$\mathcal{F}(L, p, s) = \mathcal{F}(L, (p_0, p_1 \wedge s, p_2 \wedge s, ..., p_L \wedge s, p_{L+1}), s).$$

(Idee: in keinem 'Hidden Layer' können mehr als s viele Nodes gebraucht werden.)
Damit können wir effektiv V aus Hilfslemma 2 folgendermaßen abschätzen

$$V = \prod_{l=0}^{L+1} (p_l + 1) \leq (p_0 + 1)(p_{L+1} + 1)(s+1)^L \leq 4 p_0 p_{L+1} (s+1)^L$$

und erhalten nach demselben Lemma:

$$\log(N(\delta, \mathcal{F}(L, p, s, \infty), || \cdot ||_\infty)) \leq (s+1) \log \left(2 \frac{L+1}{\delta}\, 4 p_0^2 p_{L+1}^2\, (s+1)^{2L} \right)$$

$$= (s+1) \log \left(8 \frac{L+1}{\delta}\, p_0^2 p_{L+1}^2\, (s+1)^{2L} \right)$$

Wenn wir $\delta_n = \frac{1}{n}$ setzen folgt daraus

$$\log\left(N\left(\frac{1}{n}, \mathcal{F}(L,p,s,\infty), ||\cdot||_\infty\right)\right) \leq (s+1)\log\left(8n(L+1)\,p_0^2 p_{L+1}^2\,(s+1)^{2L}\right).$$

Wir berechnen damit

$$\begin{aligned}
&\max\left(F^2\frac{18\log(N(\delta_n,\mathcal{F},||\cdot||_\infty))+76}{n\varepsilon} + 38\,\delta_n F, F^2\frac{18\log(N(\delta_n,\mathcal{F},||\cdot||_\infty))+72}{n\varepsilon} + 32\,\delta_n F\right)\\
&= F^2\frac{18\log(N(\frac{1}{n},\mathcal{F},||\cdot||_\infty))+76}{n\varepsilon} + 38\,\frac{1}{n}F\\
&\leq F^2\frac{18(s+1)\log\left(8n(L+1)\,p_0^2p_{L+1}^2\,(s+1)^{2L}\right)+76}{n\varepsilon} + 38\,\frac{1}{n}F\\
&= F^2\frac{18(s+1)\log\left(8n(L+1)\,p_0^2p_{L+1}^2\,(s+1)^{2L}\right)}{n\varepsilon} + \underbrace{\frac{76}{n\varepsilon} + 38\,\frac{1}{n}F}_{\text{fällt für } n\to\infty \text{ schneller als der Linke Term}}\\
&\leq C'\,F^2\frac{18(s+1)\log\left(8n(L+1)\,p_0^2p_{L+1}^2\,(s+1)^{2L}\right)}{n\varepsilon}\\
&= C'\,F^2\frac{18(s+1)2\log\left(\sqrt{n}p_0p_{L+1}\,(s+1)^{L}\right)}{n\varepsilon} + \underbrace{C'\,F^2\frac{18(s+1)\log\left(8(L+1)\right)}{n\varepsilon}}_{\text{fällt für } n\to\infty \text{ schneller als der Linke Term}}\\
&\leq C''\,F^2\frac{18(s+1)2\log\left(\sqrt{n}p_0p_{L+1}\,(s+1)^{L}\right)}{n\varepsilon} \leq C''\,F^2\frac{18(s+1)2\log\left(np_0p_{L+1}\,(s+1)^{L}\right)}{n\varepsilon}\\
&= \underbrace{\underbrace{\frac{36C''}{\varepsilon}}_{=:\frac{C_\varepsilon}{(1+\varepsilon)^2}}\;F^2\frac{(s+1)\log\left(np_0p_{L+1}\,(s+1)^{L}\right)}{n}}_{=\frac{\tau_{\varepsilon,n}}{(1+\varepsilon)^2}}.
\end{aligned}$$

Nun wenden wir Hilfslemma 1 an und erhalten

$$\begin{aligned}
&(1-\varepsilon)^2\Delta_n(\hat{f},f_0) - \tau_{\varepsilon,n} \leq (1-\varepsilon)^2\Delta_n(\hat{f},f_0) - \frac{\tau_{\varepsilon,n}}{(1+\varepsilon)^2} \overset{\text{HL1}}{\leq} R(\hat{f},f_0)\\
&\overset{\text{HL1}}{\leq} (1+\varepsilon)^2\left(\inf_{f\in\mathcal{F}} E[(f(X)-f_0(X))^2] + \frac{\tau_{\varepsilon,n}}{(1+\varepsilon)^2} + \Delta_n(\hat{f},f_0)\right)\\
&= (1+\varepsilon)^2\left(\inf_{f\in\mathcal{F}} E[(f(X)-f_0(X))^2] + \Delta_n(\hat{f},f_0)\right) + \tau_{\varepsilon,n}\\
&\leq (1+\varepsilon)^2\left(\inf_{f\in\mathcal{F}} ||f-f_0||_\infty^2 + \Delta_n(\hat{f},f_0)\right) + \tau_{\varepsilon,n}.
\end{aligned}$$

□