Exercise 2

Assume there exists such assignment of weights.
Then, consider 2 cases:

① $\exists\, i \in \mathbb{N}$, s.t. $w(h_i) > 0$. Then, $\exists\, \varepsilon > 0$, such that $w(h_i) > \varepsilon$. Since weights are monotonically nondecreasing, $\forall\, j > i, j \in \mathbb{N}$ it holds that $w(h_j) \geq w(h_i) > \varepsilon$.
Since the hypothesis class is infinite, there will be infinite number of weights which are greater than ε.
Since $\varepsilon > 0$, at some point the sum of weights will be greater than $\underbrace{\varepsilon + \varepsilon + \ldots + \varepsilon}_{\lceil \frac{1}{\varepsilon} \rceil \text{ times}}$ and, therefore, it will be greater than 1 and the condition $\sum_{h \in \mathcal{H}} w(h) \leq 1$ won't be satisfied.
We get a contradiction.

② $\forall\, i \in \mathbb{N}$, $w(h_i) = 0$

Note that $\forall\, i \in \mathbb{N}$, as $\delta \to 0$, $m^{uc}_{\mathcal{H}_n}(\varepsilon, \delta) \to \infty$ and $\varepsilon_n(m, \delta)$ can not be computed. Then, bound $\varepsilon_n(m, w(h_i) \cdot \delta)$ can not be minimized, since $\forall\, i \in \mathbb{N}, w(h_i) = 0$. Hence, $\mathcal{H}$ can not be nonuniformly learnt.
We get a contradiction.

Therefore, our assumption was wrong and it is impossible to assign weights to the hypotheses in $\mathcal{H}$ in a way that is described.

Exercise 4

Using Theorem 7.7 we get that:

$$\mathbb{P}_{S\sim D^m}\left[L_D(h_S) \le L_S(h_S) + \sqrt{\frac{|h_S| + \ln(2/\delta)}{2m}} \right] \ge 1-\delta$$

Since $h_S \in \arg\min_{h\in\mathcal{H}} \left[L_S(h) + \sqrt{\frac{|h| + \log(2/\delta)}{2m}} \right]$, we get that

$$L_S(h_S) + \sqrt{\frac{|h_S| + \log(2/\delta)}{2m}} \underset{\forall B}{\le} L_S(h^*_B) + \sqrt{\frac{|h^*_B| + \log(2/\delta)}{2m}} \le L_S(h^*_B) + \sqrt{\frac{B + \log(2/\delta)}{2m}}$$

Hence, we get that $\mathbb{P}_{S\sim D^m}\left[L_D(h_S) \le L_S(h^*_B) + \sqrt{\frac{B + \log(2/\delta)}{2m}} \right] \ge 1-\delta$ (1)

From Hoffding's bound (Equation 4.2), we know that:

$$D^m\left(\left\{S : \left|L_S(h^*_B) - L_D(h^*_B)\right| < \sqrt{\frac{\log(2/\delta)}{2m}}\right\}\right) \ge 1-\delta \quad (2)$$

Combining (1) and (2), we get:

$$\mathbb{P}_{S\sim D^m}\left[L_D(h_S) - L_D(h^*_B) \le \sqrt{\frac{B + \log(2/\delta)}{2m}} + \sqrt{\frac{\log(2/\delta)}{2m}} \right] \ge 1-\delta$$

bound in terms of B, δ and m

Exercise 5.5

Consider the class $\mathcal{H}_2$ of all functions from $[0,1]$ to $\{0,1\}$. Then, $\exists$ infinite set S from $[0,1]$, which is shattered by $\mathcal{H}_2$. Then, by the result of $\boxed{5.3}$, we get that for every sequence of classes $(\mathcal{H}_n : n\in\mathbb{N})$, s.t. $\mathcal{H} = \bigcup_{n\in\mathbb{N}} \mathcal{H}_n$, $\exists\, n$ for which $VCdim(\mathcal{H}_n) = \infty$. Then, $\mathcal{H}$ cannot be represented as a union of agnostic PAC-learnable hypothesis classes (result of the Fundamental Theorem of Statistical Learning). Then, by theorem 7.2, we get that $\mathcal{H}_2$ can not be nonuniformly learnt.

$\boxed{5.4}$ Define $\mathcal{H}_{1i}$ as a set of all functions which assign either 0 or 1 ~~...~~ if $x \in \{(1/2)^0, (1/2)^1, \dots (1/2)^i\}$ and assign strictly 0 to all other values from $[0,1]$.

Then, for a fixed i, $\mathcal{H}_i$ is finite and, therefore, agnostic PAC-learnable.

Let $\mathcal{H}_1 = \bigcup_{n \in \mathbb{N}} \mathcal{H}_{1n}$. Then, it is a countable union of ~~...~~ agnostic PAC-learnable hypothesis.

By theorem 7.2, this means that $\mathcal{H}_1$ is nonuniformly learnable.

Note that for every $j \in \mathbb{N}$, $\mathcal{H}_j$ shatters set $C_j = \{(1/2)^0, (1/2)^1, \dots (1/2)^j\}$. Since $\mathcal{H}_1 = \bigcup_{n \in \mathbb{N}} \mathcal{H}_{1n}$, it can shatter ~~an arbit~~ such set C_j for an arbitrary large j. Hence, $\mathcal{H}_1$ shatters sets of arbitrary large size and, therefore, has infinite VC-dimension.

Therefore, $\mathcal{H}_1$ is not PAC-learnable by Fundamental Theorem of Statistical Learning

Exercise 6

6.1 Since D is a probability distribution, $\sum_{i=1}^{\infty} D(\{x_i\}) = 1$

Then, $\forall \varepsilon > 0$, we can take M large enough, so that $\sum_{i=1}^{\infty} D(\{x_i\}) - \sum_{j=1}^{M} D(\{x_j\}) < \varepsilon$.

Hence, $\sum_{i \geq n}^{\infty} D(\{x_i\}) < \varepsilon$ for all $n > M$

Therefore, $\lim_{n\to\infty} \sum_{i \geq n} D(\{x_i\}) = 0$

6.2

~~[illegible]~~ $D(\{x \in X : D(\{x\}) < \varepsilon\})$ is smaller than ε by definition.

6.3

$$\mathbb{P}_{S\sim D^m}[\exists x_i : (D(\{x_i\}) > \eta\} \text{ and } x_i \notin S)] =$$

$$= \mathbb{P}_{S\sim D^m}[\exists x_i, i<n : (D(\{x_i\}) > \eta\} \text{ and } x_i \notin S)] =$$

$$= n \mathbb{P}_{S\sim D^m}[\exists x_i, i<n : \{D(\{x_i\}) > \eta] \leq$$

$$\leq n e^{-\eta m}$$