Exercise 1

[1] ~~All~~ First, let's consider the case where VC-dimentions of both hypothesis classes are finite.

By deffinition, VC-dimension of $\mathcal{H}'$ is the size of the largest shattered set ~~of~~ by $\mathcal{H}'$. Let's devote this set $C'$.

Since $\mathcal{H}' \subseteq \mathcal{H}$, we also get that $\mathcal{H}'_{C'} \subseteq \mathcal{H}_{C'}$.

Since $\mathcal{H}'$ shatters $C'$, $\mathcal{H}'_{C'}$ is the set of all functions from $C'$ to $\{0,1\}$. Then, $\mathcal{H}_{C'}$ is also the set of all functions from $C$ to $\{0,1\}$ (because $\mathcal{H}'_{C'} \subseteq \mathcal{H}_{C'}$).

Therefore, ~~that~~ $\mathcal{H}$ shatters $C'$ by definition.

Since $C'$ is a set that can be shattered by $\mathcal{H}$, $VCdim(\mathcal{H})$ is at least $|C'|$, which is equal to $VCdim(\mathcal{H}')$

[2] Now, consider the case where $VCdim(\mathcal{H}') = \infty$.

We have shown in [1] that every set that is shattered by $\mathcal{H}'$ is also shattered by $\mathcal{H}$. Then, $\forall$ ~~m~~ $m$ $\exists C, |C| = m$, s.t. $\mathcal{H}'$ shatters $C$ and, hence, $\mathcal{H}$ shatters $C$.

Therefore, in this case $VCdim(\mathcal{H})$ will also be infinite.

q.e.d.

Exercise 2

**1.** Let $m(k, |X|) = \min(k, |X| - k)$.

Let $C = \{c_1, \dots, c_{m(k,|X|)}\} \subset X$.

Since $k \geq \min(k, |X| - k)$, functions from $\mathcal{H}^{X}_{=k}$ can assign the value $1$ to all of the elements of $C$.

Since $|X| - k \geq \min(k, |X| - k)$, functions from $\mathcal{H}^{X}_{=k}$ can assign the value $1$ to $k$ elements of $X$ outside of $C$ and thus, assign the value $0$ to all of the elements of $C$.

Moreover, from the previous two statements it implies that functions from $\mathcal{H}^{X}_{=k}$ can assign the value $1$ to any number of elements from $C$ and, therefore, the restriction of $\mathcal{H}^{X}_{=k}$ to $C$ is the set of all functions from $C$ to $\{0, 1\}$.

Therefore, $\mathcal{H}^{X}_{=k}$ shatters $C$ by definition.

Now, consider an arbitrary sample $C'$ from $X$ of the size $m(k, |X|) + 1$. Now, if $k \leq |X| - k$, then $k = \min(k, |X| - k)$ and, therefore, $k < m(k, |X|) + 1$. Then, no $h \in \mathcal{H}^{X}_{=k}$ can account for the labeling $\underbrace{(1, 1, \dots 1)}_{m(k,|X|)+1}$ and hence $C'$ is not shattered by $\mathcal{H}^{X}_{=k}$.

If $k \geq |X| - k$, then $|X| - k < \min(k, |X| - k) + 1$ and no $h \in \mathcal{H}^{X}_{=k}$ can account for the labeling $\underbrace{(0, \dots, 0)}_{m(k,|X|)+1}$. Hence $C'$ is not shattered by $\mathcal{H}^{X}_{=k}$.

Therefore, $VCdim(\mathcal{H}^{X}_{=k}) = \min(k, |X| - k)$

$\boxed{2}$ Let $\mathcal{H}_1 = \{h \in \{0,1\}^{\mathcal{X}} : |\{x : h(x) = 1\}| \leq k\}$ and

$\mathcal{H}_0 = \{h \in \{0,1\}^{\mathcal{X}} : |\{x : h(x) = 0\}| \leq k\}$.

Consider 2 cases:

**Case 1: $|\mathcal{X}| \leq 2k+1$**

Then, every labeling of every set $C \subseteq \mathcal{X}$ will either have at most $k$ $1^s$ or at most $k$ $0^s$, since the largest possible size for $C$ is $|\mathcal{X}|$ and $|\mathcal{X}| \leq 2k+1$ (i.e. we can't have $k+1$ $0^s$ and $k+1$ $1^s$ at the same time in our labeling)

Then, every ~~labeling~~ possible labeling will correspond to either some function $h_1 \in \mathcal{H}_1$ or to some function $h_0 \in \mathcal{H}_0$.

Since $\mathcal{H}_{at\text{-}most\text{-}k} = \mathcal{H}_1 \cup \mathcal{H}_0$, we get that $\mathcal{H}_{at\text{-}most\text{-}k}$ shatters every set $C \subseteq \mathcal{X}$.

Therefore, in this case, $VCdim(\mathcal{H}_{at\text{-}most\text{-}k}) = |\mathcal{X}|$, because the largest set $C \subseteq \mathcal{X}$ is $\mathcal{X}$ itself.

**Case 2: $|\mathcal{X}| > 2k+1$** (then, $|\mathcal{X}| \geq 2k+2$, since $|\mathcal{X}| \in \mathbb{N}$)

Consider an arbitrary set $C$ of size $2k+1$. Then, for every function from $C$ to $\{0,1\}$, it will either label at most $k$ elements of $C$ as $0$ or it will label at most $k$ elements of $C$ as $1$.

(since otherwise it will label $k+1$ elements as $1$ and $k+1$ elements as $0$, which is not possible for the set of size $2k+1$)

Therefore, $\mathcal{H}_0 \cup \mathcal{H}_1$ is the set of all functions from $C$ to $\{0,1\}$, hence $\mathcal{H}_{at\text{-}most\text{-}k} = \mathcal{H}_0 \cup \mathcal{H}_1$ shatters $C$ by definition.

Now, consider an arbitrary set $C'$ of size $2k+2$. Consider labeling ~~...~~ $(\underbrace{1, \ldots, 1}_{k+1}, \underbrace{0, \ldots, 0}_{k+1})$. Then, no $h \in \mathcal{H}_{at\text{-}most\text{-}k}$ can account for this labeling and, therefore, $C'$ is not shattered by $\mathcal{H}_{at\text{-}most\text{-}k}$

Therefore, in this case, $VCdim(\mathcal{H}_{at\text{-}most\text{-}k}) = 2k+1$.

Combining cases 1 and 2, we get the generalized answer, which is:

$$VCdim(\mathcal{H}_{at\text{-}most\text{-}k}) = \min(|X|, 2k+1)$$

## Exercise 9

First, take the set $C = \{1,2,3\}$. Then, for every labeling of this set, we have the corresponding $h \in \mathcal{H}$:

① $(-1,-1,-1) \rightarrow h_{0,4,-1}(x)$ ⑤ $(1,1,1) \rightarrow h_{0,4,1}(x)$

② $(-1,-1,1) \rightarrow h_{0,2,-1}(x)$ ⑥ $(1,1,-1) \rightarrow h_{0,2,1}(x)$

③ $(-1,1,-1) \rightarrow h_{1.5,2,1}(x)$ ⑦ $(1,-1,1) \rightarrow h_{1.5,2,-1}(x)$

④ $(1,-1,-1) \rightarrow h_{2,3,-1}(x)$ ⑧ $(-1,1,1) \rightarrow h_{2,3,1}(x)$

Therefore, $\mathcal{H}$ shatters $C$.

Now, consider an arbitrary set $C = \{c_1, c_2, c_3, c_4\}$. Without loss of generality, assume that $c_1 \leq c_2 \leq c_3 \leq c_4$. Then, the labeling $(-1,1,-1,1)$ cannot be obtained by a signed interval and, therefore, the VC dimension of the class of signed intervals is 3.

## Exercise 10

**2** Consider $\mathcal{H}$ that is PAC learnable.

Assume $VCdim(\mathcal{H})$ is infinite.

Then, $\forall$ sample sizes $m$, $\exists$ shattered $\overset{(\text{by } \mathcal{H})}{}$ set of size $2m$.

Let's fix an arbitrary sample size $m$ and consider a shattered set $C$ of size $2m$.

Then, by COROLARY 6.4, for any learning algorithm $A$, $\exists$ distribution $\mathcal{D}$ over $\mathcal{X} \times \{0,1\}$, s.t.:

$$\mathbb{P}_{S \sim \mathcal{D}^m} [L_{\mathcal{D}}(A(S)) \geq 1/8] \geq 1/7$$

This means that $\forall$ sample sizes $m$, $\forall$ learning algorithm $A$, $\exists$ distribution $\mathcal{D}$ over $\mathcal{X} \times \{0,1\}$ and labeling function $f: \mathcal{X} \to \{0,1\}$, s.t. $\mathbb{P}_{S \sim \mathcal{D}^m} [L_{\mathcal{D}}(A(S)) \geq 1/8] \geq \frac{1}{7}$

Hence, $\mathcal{H}$ can not be PAC learnable, which is a contradiction.

Therefore, our assumption was wrong and $VCdim(\mathcal{H}) < \infty$

q.e.d.