Exercise 1

Let $\mathcal{H} = \{h_1, h_2, h_3, h_4\}$, where:

$$h_1(x) = \begin{cases} 1, & \text{if } x=1 \\ 1, & \text{if } x=2 \\ 0, & \text{if } x=3 \\ 1, & \text{if } x=4 \end{cases} \qquad h_2(x) = \begin{cases} 0, & \text{if } x=1 \\ 0, & \text{if } x=2 \\ 0, & \text{if } x=3 \\ 1, & \text{if } x=4 \end{cases}$$

$$h_3(x) = \begin{cases} 0 & \text{if } x=1 \\ 1 & \text{if } x=2 \\ 1 & \text{if } x=3 \\ 1 & \text{if } x=4 \end{cases} \qquad h_4(x) = \begin{cases} 0, & \text{if } x=1 \\ 1, & \text{if } x=2 \\ 0, & \text{if } x=3 \\ 1, & \text{if } x=4 \end{cases}$$

and let sequence S be $\{(1,0),(2,1),(3,0),(4,1)\}$

Now, let's implement Consistent Algorithm:

$V_1 = \mathcal{H}$

for $t=1$: choose $h_1 \in V_1$, predict $p_1 = h_1(1) = 1 \neq h^*(1)$

$$V_2 = \{h \in V_1 : \underset{\uparrow \atop 1}{h(x_1)} = \underset{\uparrow \atop 0}{y_1}\} = \{h_2, h_3, h_4\}$$

For $t=2$: choose $h_2 \in V_2$, predict $p_2 = h_2(2) = 0 \neq h^*(2)$

$$V_3 = \{h \in V_2 : h(2) = 1\} = \{h_3, h_4\}$$

for $t=3$: choose $h_3 \in V_3$, predict $p_3 = h_3(3) = 1 \neq h^*(3)$

$$V_4 = \{h \in V_3 : h(3) = 0\} = \{h_4\}$$

for $t=4$, the only possible prediction is h_4, which is a correct one.

As we can see, for this hypothesis class $\mathcal{H}$ and sequence S, Consistent makes $3 = 4-1 = |\mathcal{H}|-1$ misstakes.

$\boxed{\text{Exercise 2}}$ Let $\mathcal{H} = \{h_1, h_2, h_3, h_4\}$, where:

$$h_1(x) = \begin{cases} 1, \text{if } x=1 \\ 0, \text{if } x=2 \\ 1, \text{if } x=3 \\ 1, \text{if } x=4 \end{cases} \quad h_2(x) = \begin{cases} 1, \text{if } x=1 \\ 0, \text{if } x=2 \\ 1, \text{if } x=3 \\ 0, \text{if } x=4 \end{cases} \quad h_3(x) = \begin{cases} 0, \text{if } x=1 \\ 1, \text{if } x=2 \\ 1, \text{if } x=3 \\ 1, \text{if } x=4 \end{cases} \quad h_4(x) = \begin{cases} 0, \text{if } x=1 \\ 0, \text{if } x=2 \\ 1, \text{if } x=3 \\ 1, \text{if } x=4 \end{cases}$$

and let sequence S be $\{(1,0), (2,0), (3,1), (4,1)\}$.

Apply Halving algorithm:

$V_1 = \mathcal{H}$

for $t = 1$:

$p_1 = \underset{r \in \{0,1\}}{\operatorname{argmax}} |\{h \in V_1 : h(1) = r\}| = 1$ (since halving predicts 1 in case of a tie)

$h^*(1) = 0$, so the prediction is wrong

$V_2 = \{h \in V_1 : h(1) = 0\} = \{h_3, h_4\}$.

for $t = 2$:

$p_2 = \underset{r \in \{0,1\}}{\operatorname{argmax}} |\{h \in V_2 : h(2) = 0\}| = 1$ (because of a tie)

$\nparallel$

$h^*(2) = 0$, so the prediction is wrong

$V_3 = \{h_4\}$

~~therefore~~: The only remaining hypothesis is the correct one, therefore halving doesn't make any other misstakes.

As we can see, for hypothesis class $\mathcal{H}$ and sequence S, Halving makes $2 = \log_2(4) = \log_2(|\mathcal{H}|)$ misstakes.

Exercise 3

labeling fcn of the sequence S

Lower Bound Fix some $d \geq 2$. Consider $h^* = h_d$. Then, the last version space V_d should only have the one and only correct hypotheses. We know that $|V_{t+1}| \leq \frac{1}{2}|V_t|$ for Halving algorithm, hence $|V_{d-1}| \geq 2|V_d| = 2$. Since we $h^* = h_d$, $h_d^*(x_{d-1}) = 0$. Therefore, there is a hypothesis in V_{d-1} which incorrectly predicted label for x_{d-1} to be 0. Since the only such hypothesis is h_{d-1}, we get that $V_{d-1} = \{h_{d-1}, h_d\}$. Then we have that:

$$1 = |V_{d-1}^1| \geq |V_{d-1}^0| = 1$$

and, hence, $p_{d-1} = 1$, which is an incorrect prediction.

Therefore, $M_{HALVING}(\mathcal{H})$ is at least 1.

Upper Bound Fix some h^* and sequence S.

Consider 2 types of mistakes:

1. $p_t = 1$, but $h^*(x_t) = 0$. Then, this means that $|V_t^1| \geq |V_t^0|$ and, since $|V_i^1| = 1$ for all $i \in [d]$, $|V_t^0| \leq 1$. Hence, $|V_t| = |V_t^0| + |V_t^1| \leq 2$. Since $|V_{t-1}| \geq 2|V_t|$ and this type of mistake doesn't happen for $|V_i| > 2$, we get that this type of mistake have not occured for all $i \in [t-1]$. Let $\hat{t}$ be such that $h_{\hat{t}} = h^*$. Then, for all $n > t^*$, V_n has only the correct hypothesis and thus p_n doesn't make any mistakes. Since p_t makes a mistake, $t < \hat{t}$ and V_t must contain at least one incorrect hypothesis. Then, $|V_t| \geq 2$. Since it is also ≤ 2, we get that $|V_t| = 2$. We know that one of the two hypotheses in V_t is $h_{\hat{t}}$,

and it is the one which predicts ~~wrong~~ the label for x_t to be 0. Since $|V_t^1| = 1$, the second hypothesis in V_t is h_t. This means that $\hat{t} = t+1$ and, therefore this mistake is the last one algorithm makes (V_{t+1} will only have the correct hypothesis)
We also get that no mistakes have happened in the previous $t-1$ steps. Therefore, we have shown that when the mistake of this type happens, it is the only mistake that the algorithm makes

② $p_t = 0$, but $h^*(x_t) = 1$. Since there is only one x_t, s.t. $h^*(x_t) = 1$, this misstake can only happen once.

Those are the only two types of mistakes. Since there can be at most one mistake of each type and the mistakes of different types are mutually exclusive, we get that $M_{HALVING}(\mathcal{H}) \leq 1$

Hence, combining upper and lower bounds, we get that $M_{HALVING}(\mathcal{H}) = 1$