Exercise 1

Equation 5.2 says that $\underset{S\sim D^m}{\mathbb{E}}[L_D(A'(S))] \geq \frac{1}{4}$

We want to show that it implies that $\mathbb{P}[L_D(A(S)) \geq \frac{1}{8}] \geq \frac{1}{2}$

Since $L_D(A(S))$ is an expectation of the 0-1 loss function, it takes values in $[0,1]$.

Then, we can use ~~theorem~~ lemma B.1:

$$P[L_D(A(S)) \geq \frac{1}{8}] \geq P[L_D(A(S)) > \frac{1}{8}] \geq \frac{\mathbb{E}_{S\sim D^m}[L_D(A(S))] - 1/8}{1 - 1/8} \geq$$

$$\geq \frac{1/4 - 1/8}{7/8} = \frac{2-1}{7} = 1/7$$

Hence, ~~$\mathbb{P}_{L_D}(A(S))$~~ $\mathbb{P}[L_D(A(S)) \geq \frac{1}{8}] \geq 1/7$

q. e. d.

Exercise 2

1. The hypothesis class of the first algorithm is more restrictive than the hypothesis class of the second algorithm and, therefore the first algorithm will be likely to have more inductive bias.
Therefore, the first algorithm will be prone to a higher approximation error than the second algorithm.
Since the second algorithm takes more parameters into consideration, its hypothesis class is ~~more~~ likely to have a higher complexity and the algorithm has a higher chance of overfitting compared to the first algorithm.
Therefore the second algorithm is likely to have a higher estimation error.

2. Since approximation error does not depend on the sample size, but the estimation error can be reduced by a larger sample, higher number of available labeled training & samples will ~~the~~ work in favour of the second algorithm.