1. Exercise 2.8

   (a) We know that $\overline{g}$ is the average function of many different hypothesis $g_1,\ g_2, ...,\ g_n$ of different data sets. H represents the hypotheses set, where each hypothesis in H is dependent on their respective data set. If the hypothesis in H are in linear combination, then the average of the hypotheses in H should also be a linear combination, proving that $\overline{g} \in H$ ■

   (b) We can imagine a model with two hypotheses, one that will label all datapoints as $+1$ and another as -1, then the average of those two hypotheses will be 0, which is not in the hypothesis set, therefore $\overline{g} \notin H$ ■

   (c) (b) is a binary classification and $\overline{g}$ is not a binary function since $0 \notin \{+1, -1\}$. Often, with more hypotheses, the average will be a number between $\{-1, 1\}$, it will be unlikely that they are $+1$ or -1

2. Problem 2.14

   (a) Given that the d_{vc} is finite, we know that the hypothesis can shatter any data set of size d_{vc}. We can claim that $K(d_{vc}) \geq d_{vc}(H)$, since $K(d_{vc})$ assumes that every K hypothesis can shatter the maximum number of points, d_{vc}. Then $K(d_{vc}+1) > K(d_{vc}) \geq d_{vc}(H)$, by transitivity, $K(d_{vc}+1) > d_{vc}(H)$ ■

   (b) Given that the hypothesis has a finite VC dimension, then we have a breakpoint such that $2^\ell > \ell^{d_{vc}} + 1$ If we have K hypotheses, then $m_H(\ell) \leq K(\ell^{d_{vc}} + 1)$. By inspection, $2K\ell^{d_{vc}} > K(\ell^{d_{vc}} + 1)$. So here we know that $2^\ell > 2K\ell d^{vc} > K(\ell^{d_{vc}} + 1)$, by transitivity, $2^\ell > K(\ell^{d_{vc}} + 1)$, which means $2^\ell > m_H(\ell)$, if this is true, then it means that there are no hypothesis can shatter 2^ℓ points, which means that $d_{vc} \leq \ell$ ■

   (c) **TO DO**

3. Problem 2.15

   (a)