1. Exercise 3.4

   (a) we know $y = w^{*T}x + \epsilon$ and $H = X(X^TX)^-1X^T$ from (3.6), and we know $\hat{y} = Hy$ by definition, we want to prove $\hat{y} = Xw^* + H\epsilon$

$$\begin{aligned}\hat{y} &= H(w^*X + \epsilon)\\ &= X(X^TX)^{-1}X^T(w^*X + \epsilon)\\ &= X(X^TX)^{-1}X^Tw^*X + X(X^TX)^{-1}X^T\epsilon\\ &= w^*X + H\epsilon\end{aligned}$$

   (b) for $\hat{y} - y$, we have

$$\begin{aligned}\hat{y} - y &= w^*X + H\epsilon - (w^* + \epsilon)\\ &= H\epsilon - \epsilon\\ &= \epsilon(H - I)\end{aligned}$$

   where I denotes the identity matrix

   (c) let $E_{in}(w) = \frac{1}{N}||\hat{y} - y||^2$

$$\begin{aligned}E_{in}(w) &= \frac{1}{N}||\epsilon(H - I)||^2\\ &= \frac{1}{N}\epsilon^T(H - I)^T\epsilon(H - I)\end{aligned}$$

   We know $H - I$ is symmetric, so $(H - I)^T = (H - I)$

$$\begin{aligned}E_{in}(w) &= \frac{1}{N}\epsilon^T\epsilon(H - I)^2\\ &= \frac{1}{N}\epsilon^T\epsilon(I - H)^2\end{aligned}$$

   (d) We know

$$\begin{aligned}E_D[E_{in}(w_{lin})] &= E_D[\frac{1}{N}(\epsilon^T\epsilon(I - H))]\\ &= \frac{1}{N}(E_D[\epsilon^T\epsilon] - E_D[\epsilon^T\epsilon H])\end{aligned}$$

   Given that ϵ is a noise term with zero mean and σ^2 variance. The variance of each noise component ϵ is σ^2, so

$$\begin{aligned} E_D[E_{in}(w_{lin})] &= \frac{1}{N}(N\sigma^2 - E_D[\epsilon^T \epsilon H]) \\ &= \sigma^2 - \frac{1}{N} E_D[\epsilon^T \epsilon H] \end{aligned}$$

finish