1 The Ridge Regression Estimator

See my notes on ridge regression.

2 Entropy and Information

Part (a)

$$\Pr(X = x) = (1-b)^{(x-1)} \cdot b$$

$$\begin{aligned}
H(X) &= -\sum_{x=1}^{\infty}((1-b)^{(x-1)} \cdot b)((x-1)\log(1-b) + \log b) \\
&= -b\sum_{x=1}^{\infty}\left(\log(1-b)\cdot(x-1)(1-b)^{(x-1)}\right) + \left(\log b \cdot (1-b)^{(x-1)}\right) \\
&= -b\left(\log(1-b)\cdot\frac{1-b}{b^2} + \log b \cdot \frac{1}{b}\right)
\end{aligned}$$

Part (b)

If $b = 1/2$, $H(X) = 2$. For every round k, we ask $X = k$? This implies the answer was 0 in all previous rounds ($1, \ldots, k-1$). Note that the probability of the number of questions that we need to ask equals k is $(\frac{1}{2})^k$. The expected number of questions to ask is

$$\sum_{k=1}^{\infty} k \cdot (\frac{1}{2})^k = 2 = H(X)$$

3 Decision Trees

Part (a)

False. The statement would be true if the splits were allowed to form complex boundaries (not binary and linear).

Part (b)

False. XOR example.

Part(d)

Deep $\Rightarrow$ High variance & Low bias (overfit).
Shallow $\Rightarrow$ Low variance & High bias (underfit).