Solutions by **Andrew Lys** andrewlys(at)u.e.

1. **Feature Selection.**

(a) i. Let $k = 1$. Then since x_1 and x_{100} are uncorrelated, we simply pick our feature as the feature which contributes more to the signal, namely x_100. Our predictor is then

$$h_w(x) = ax_{100}$$

And we pick $a = \frac{3}{\sqrt{10}}$. Our error is then:

$$L_D(h_w) = E[x_1^2/10] = \frac{1}{10}\operatorname{Var}(x_1) = \frac{1}{10}$$

For $k = 2$ and above, we simply pick x_1 and x_{100} as our features, $w = \frac{3}{\sqrt{10}}e_{100} + \frac{1}{\sqrt{10}}e_1$. We get zero loss.

2. **Boosting as Coordinate Descent.**