

1. What are the sizes of the normalized training and testing data sets and the associated labels?
2. Suppose you have applied the GD algorithm to minimize  $f(\hat{\mathbf{w}})$  in (E4.2) and obtained a minimizer  $\{\mathbf{w}^*, b^*\}$ , what are the sizes of  $\mathbf{w}^*$  and  $b^*$ ?
3. What are sizes of  $\hat{\mathbf{x}}_n$  and  $\mathbf{x}_n$ ?