Exercise 1

(a) Many solutions are possible here. The easiest way is to split the data $\mathcal{D}_n$ into a test set and a training set. See bottom of page 11 of the lecture notes. Another solution is to use K-fold cross validation as described on page 13 of the lecture notes. More concretely, one solution is to say that for some $n_1 < n$, $i = 1, \dots, n_1$ is the training set and $i = n_1 + 1, \dots, n$ is the test set. Then the generalization error of an algorithm $\hat{m}$ can be estimated via

$$T(\hat{m}) = \sum_{i=1}^{n_1} L(\hat{m}(\{X_j, Y_j\}_{j=n_1+1,\dots,n})(X_i), Y_i)$$

(b) Let $T(\hat{m})$ be the estimated generalisation error described in part (a). Note that T will necessarily depend on $\mathcal{D}_n$. Then

$$\text{Bias} = \mathbb{E}_{\mathcal{D}_n}[T(\hat{m})] - \mathbb{E}_{X,Y,\mathcal{D}_n}[L(Y, \hat{m}(\mathcal{D}_n)(X))], \quad \text{Variance} = \text{Var}_{\mathcal{D}_n}[T(\hat{m})].$$

(c) In the procedure described in (a) the data is split into training and testing (possibly multiple times). Bias can be reduced by increasing the training size.