Exercise 1

(a) Given iid data $\mathcal{D}_n = (X_i, Y_i)_{i=1,\ldots,n}$ and an algorithm $\hat{m}$, describe one way of how you can estimate the unconditional generalization error

$$\mathbb{E}_{(X,Y,\mathcal{D}_n)}[L(Y, \hat{m}(\mathcal{D}_n)(X))],$$

where (X, Y) are independent copies of (X_i, Y_i)

(b) The procedure you described in (a) will have a bias and a variance. Write down the precise mathematical definition of the bias and variance considered here.

(c) Explain how the procedure in (a) can be altered to reduce reduce bias.