$$\mathbb{E} R_n^{(D)}(g_n) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\, \mathbb{1}_{g_{n-1}(X_i, D_{n,i}) \neq Y_i}$$ (linearity of expectation)

$$= \mathbb{E}\, \mathbb{1}_{g_{n-1}(X, D_{n-1}) \neq Y}$$ (by identical distribution)

$$= \mathbb{E} R(g_{n-1}).$$

From the previous exercise we see that

$$R_n^{(D)}(g_n) \leq \frac{\|w_0 - w_*\|^2}{n},$$ because by removing a point on which the algorithm does not update, the classifier does not change and that point is correctly classified.

For example, if $w_0 = 0$, then we get

$$\mathbb{E} R(g_{n-1}) \leq \frac{\mathbb{E}\|w_*\|^2}{n}$$