The following lower bound on the expected excess risk holds for binary classification.

Theorem. *Let $n \in \mathbb{N}$ be the training set size and $\mathcal{G}$ be a set of prediction functions of VC-dimension $V \leq 4n$. For any estimator $\hat{g}$, there exists a probability distribution for which*

$$\mathbb{E}R(\hat{g}) - \inf_{g \in \mathcal{G}} R(g) \geq \frac{1}{8}\sqrt{\frac{V}{n}}.$$

The point of this note is to put forward the (mild but) omitted assumption $V \leq 4n$, and also to mention that the result holds for $V = 1$ (and trivially for $V = 0$). There is no need to correct the proof of the Annals of Statistics version of the paper. We still choose the $(V, 1/V, (1-2L)^2)$-hypercube with L such that $1 - 2L = \frac{1}{2}\sqrt{V/n}$. The condition $V \leq 4n$, which was unfortunately omitted in the published version, comes from the constraint $1 - 2L \leq 1$. Note also that for $V > 4n$, by choosing $L = 0$, we have

$$\mathbb{E}R(\hat{g}) - \inf_{g \in \mathcal{G}} R(g) \geq \frac{1 - \sqrt{n/V}}{2} \geq \frac{1}{4}.$$