Our algorithm outputs a stochastic classifier, $h_{\mathcal{A}}$ that is a convex combination of other hypotheses in $\mathcal{H}$. We denote the class of functions which are convex combinations of functions in $\mathcal{H}$ as $conv(\mathcal{H})$.

Lemma 1. *The Rademacher complexity defined over a convex combination of distributions is linear with respect to the sample distribution.*

Proof.

$$\begin{aligned}
\mathcal{R}_{\mathcal{D}}(\mathcal{F}) &= \mathbb{E}_{S\sim\mathcal{D}}\left[\hat{\mathcal{R}}_S(\mathcal{F})\right] \\
&= \sum_S \mathbb{P}_{\mathcal{D}}\left[S\right]\left(\hat{\mathcal{R}}_S(\mathcal{F})\right) \\
&= \sum_S \left(\alpha\mathbb{P}_{\mathcal{D}_A}\left[S\right] + (1-\alpha)\mathbb{P}_{\mathcal{D}_B}\left[S\right]\right)\left(\hat{\mathcal{R}}_S(\mathcal{F})\right) \\
&= \sum_S \alpha\mathbb{P}_{\mathcal{D}_A}\left[S\right]\left(\hat{\mathcal{R}}_S(\mathcal{F})\right) + (1-\alpha)\mathbb{P}_{\mathcal{D}_B}\left[S\right]\left(\hat{\mathcal{R}}_S(\mathcal{F})\right) \\
&= \sum_S \alpha\mathbb{P}_{\mathcal{D}_A}\left[S\right]\left(\hat{\mathcal{R}}_S(\mathcal{F})\right) + \sum_S (1-\alpha)\mathbb{P}_{\mathcal{D}_B}\left[S\right]\left(\hat{\mathcal{R}}_S(\mathcal{F})\right) \\
&= \alpha\sum_S \mathbb{P}_{\mathcal{D}_A}\left[S\right]\left(\hat{\mathcal{R}}_S(\mathcal{F})\right) + (1-\alpha)\sum_S \mathbb{P}_{\mathcal{D}_B}\left[S\right]\left(\hat{\mathcal{R}}_S(\mathcal{F})\right) \\
&= \alpha\mathcal{R}_{\mathcal{D}_A}(\mathcal{F}) + (1-\alpha)\mathcal{R}_{\mathcal{D}_B}(\mathcal{F}) \qquad \square
\end{aligned}$$

Lemma 2. *Let $\mathcal{F}' = conv(\mathcal{F})$. Then $\mathcal{R}_{\mathcal{D}}(\mathcal{F}') = \mathcal{R}_{\mathcal{D}}(\mathcal{F})$.*

Proof. Find citation. $\square$