Supplementary Materials: Appendix

7 MISSING PROOFS

Proof. **(Proposition 1)** We will define a set of n labelers with error rates ϵ, a distribution $\mathcal{D}$ over the domain $\mathcal{X}$, and a binary classification y of $\mathcal{X}$ that achieve an error rate for f^* equal to $2\epsilon n/(n+1)$. Number the labelers from 0 to $n-1$. The domain $\mathcal{X}$ can be partitioned in $n+1$ subsets A and $B_0, \ldots, B_{n-1}$ so that $\mathbb{P}_{x\sim\mathcal{D}}(A) = 1 - 2n\epsilon/(n+1)$ and $\mathbb{P}_{x\sim\mathcal{D}}(B_i) = 2\epsilon/(n+1)$ for $i = 0, \ldots, n-1$. For each $x \in \mathcal{D}$, labeler j outputs 1 if and only if $x \in B_i$, for $i \in \{j \mod n, j+1 \mod n, \ldots, j+(n-1)/2 \mod n\}$. The labelers output 0 for each $x \in A$. For each $x \in A$, we have that $y(x) = 0$ hence all labelers are correct, and the function f^* clearly agrees with the output of the labelers. For each $x \in B_0 \cup \ldots \cup B_{n-1}$, we have that $y(x) = 0$. By the definitions of the subsets $B_0, \ldots, B_{n-1}$, we can observe that for each subset B_i there are exactly $(n+1)/2$ labelers that output 1 and $(n-1)/2$ labelers that output 0. Hence, for all these subsets, the function f^* returns 1. This implies that for each region B_i, the labelers that output 1 are wrong. Since each labeler outputs 1 in $(n+1)/2$ of these subsets, each labeler is wrong in a subset of the domain that has probability $\frac{2\epsilon}{n+1} \cdot \frac{n+1}{2} = \epsilon$. Also, for each $x \in B_0 \cup \ldots \cup B_{n-1}$, the function f^* fails to provide the correct answer, therefore f^* is incorrect with probability $2\epsilon n/(n+1)$. This concludes the proof. ☐

Proof. **(Proposition 2)** Suppose (w.l.g.) that $\epsilon_1 \leq \ldots \leq \epsilon_n$. For $i = 1, \ldots, n$, let W_i be the subset of the domain where the i-th labeler is wrong, i.e. $W_i = \{x \in \mathcal{X} : \ell_i(x) \neq y(x)\}$. We choose the labelers in a way such that $W_1 \subseteq W_2 \subseteq \ldots \subseteq W_n$. For any $i = 1, \ldots, n$, the labelers from i to n are wrong in W_i, hence $n-i+1$ labelers are wrong in W_i. Hence, the majority vote is wrong for any $x \in W_{(n+1)/2}$, and the statement follows as $\mathbb{P}_{x\sim\mathcal{D}}(x \in W_{(n+1)/2}) = \epsilon_{(n+1)/2}$ ☐

Proof. **(Proposition 3).** We will prove the statement by showing that the left side of the equality is both greater or equal and less or equal than the right side. It is straightforward to see that $\max_{S\in\mathcal{S}(\vec{\epsilon})} \min_{f\in\mathcal{F}} \varepsilon(f \circ \vec{\boldsymbol{\ell}}_S) \leq \min\{\epsilon_1, \ldots, \epsilon_n\}$, as we can just copy the output of the most accurate labeler.

We now show the other inequality. Consider a set of function $S = \{\ell_1, \ldots, \ell_n\}$, and without loss of generality assume that the labelers are in non-decreasing order based on their error rates. For $i = 1, \ldots, n$, let $W_i = \{x \in \mathcal{X} : \ell_i(x) \neq y(x)\}$ be the subset of the domain where the labeler i is wrong. We choose S so that the following holds $W_1 \subseteq \ldots \subseteq W_n$. Moreover, let $y(x) = 1$ if and only if $x \notin W_1$. Let $\vec{\mathbf{1}}_n$ be a vector of n bits all set to 1. We have that $\mathbb{P}_{x\sim\mathcal{D}}(y(x) = 1 \wedge \vec{\boldsymbol{\ell}}_S(x) = \vec{\mathbf{1}}_n) = \mathbb{P}_{x\sim\mathcal{D}}(x \in (\mathcal{X} - W_n)) = 1 - \epsilon_n \geq 1/2$, and that $\mathbb{P}_{x\sim\mathcal{D}}(y(x) = 0 \wedge \vec{\boldsymbol{\ell}}_S(x) = \vec{\mathbf{1}}_n) = \mathbb{P}_{x\sim\mathcal{D}}(x \in W_1) = \epsilon_1$. Hence, any function in $\mathcal{F}$ that maps the vector $\vec{\mathbf{1}}_n$ to 0 has error at least $1 - \epsilon_n$, and any function that maps the same vector to 1 has error at least ϵ_1. Therefore, we have shown that there exists a scenario where any function in $\mathcal{F}$ has error at least $\min\{\epsilon_1, 1 - \epsilon_n\} = \epsilon_1$. This concludes the proof. ☐

Proof. **(Proposition 5)** For convenience, renumber the labelers to $1, 2$ and 3. Observe that the linear program (5) has 7 equality constraints and 8 variables. If we fix the value of the variable $p_{000} = x$, these constraints impose a unique solution for the other variables. In particular, we have that:

$$
\begin{aligned}
p_{111} &= 1 - x - \frac{\mathrm{D}_{1,2} + \mathrm{D}_{1,3} + \mathrm{D}_{2,3}}{2} \\
p_{011} &= (\mathrm{D}_{1,2} + \mathrm{D}_{1,3} - \epsilon_2 - \epsilon_3)/2 + x \\
p_{101} &= (\mathrm{D}_{1,2} + \mathrm{D}_{2,3} - \epsilon_1 - \epsilon_3)/2 + x \\
p_{110} &= (\mathrm{D}_{1,3} + \mathrm{D}_{2,3} - \epsilon_1 - \epsilon_2)/2 + x \\
p_{100} &= (\epsilon_2 + \epsilon_3 - \mathrm{D}_{2,3})/2 - x \\
p_{010} &= (\epsilon_1 + \epsilon_3 - \mathrm{D}_{1,3})/2 - x \\
p_{001} &= (\epsilon_1 + \epsilon_2 - \mathrm{D}_{1,2})/2 - x
\end{aligned}
$$

With this assignment, the objective function of (5) is equal to $\sum_{i \neq j}(\epsilon_i + \epsilon_j - \mathrm{D}_{i,j})/2 - 2x$. Therefore, the objective function is maximized by picking the least x that yields a feasible solution. The assignment above is feasible if and only if each variable is ≥ 0. By setting these constraints, we obtain the following lower-bounds for x:

$$
\begin{aligned}
p_{000} \geq 0 &\iff x \geq 0 \\
p_{100} \geq 0 &\iff x \geq (\epsilon_2 + \epsilon_3 - \mathrm{D}_{2,3})/2 \\
p_{010} \geq 0 &\iff x \geq (\epsilon_1 + \epsilon_3 - \mathrm{D}_{1,3})/2 \\
p_{001} \geq 0 &\iff x \geq (\epsilon_1 + \epsilon_2 - \mathrm{D}_{1,2})/2
\end{aligned}
$$

These lower-bounds on x can be equivalently stated as $x \geq q$. By setting $x = q$, we conclude the proof. $\square$

8 ALGORITHM OF SECTION 4.3

Here, we report the pseudocode of the algorithm presented in Section 4.3. The pseudocode uses the method of Section 4.1 to compute the worst-case error of set of labelers. Alternatively, the method of Section 4.2 can also be used.

Algorithm 1 `GreedySelection`$(\vec{\epsilon}, \mathbf{D})$

best-eps $\leftarrow 0$
best-$I \leftarrow \emptyset$
for $i = 1, \ldots, n$ **do**
    $I \leftarrow \{i\}$
    curr-eps $\leftarrow \epsilon_i$
    while $\min_{P \subseteq [n]/I : |P|=2} \max_{S \in \mathcal{S}(\vec{\epsilon}, \mathbf{D})} \varepsilon(\lambda_{\mathcal{I} \cup \mathcal{P}} \circ \vec{\boldsymbol{\ell}}_S) <$ curr-eps **do**
        $P \leftarrow \operatorname{argmin}_{P \subseteq [n]/I : |P|=2} \max_{S \in \mathcal{S}(\vec{\epsilon}, \mathbf{D})} \varepsilon(\lambda_{\mathcal{I} \cup \mathcal{P}} \circ \vec{\boldsymbol{\ell}}_S)$
        $I \leftarrow P \cup I$
        curr-eps $\leftarrow \max_{S \in \mathcal{S}(\vec{\epsilon}, \mathbf{D})} \varepsilon(\lambda_{\mathcal{I}} \circ \vec{\boldsymbol{\ell}}_S)$
    end while
    if curr-eps $<$ best-eps **then**
        best-eps $\leftarrow$ curr-eps
        best-$I \leftarrow I$
    end if
end for
return (best-I,best-eps)

9 ADDITIONAL FIGURES

We provide the additional figures for other groups of AwA2 tasks as we vary the amount of labeled data used to make estimates of the labeler accuracies.

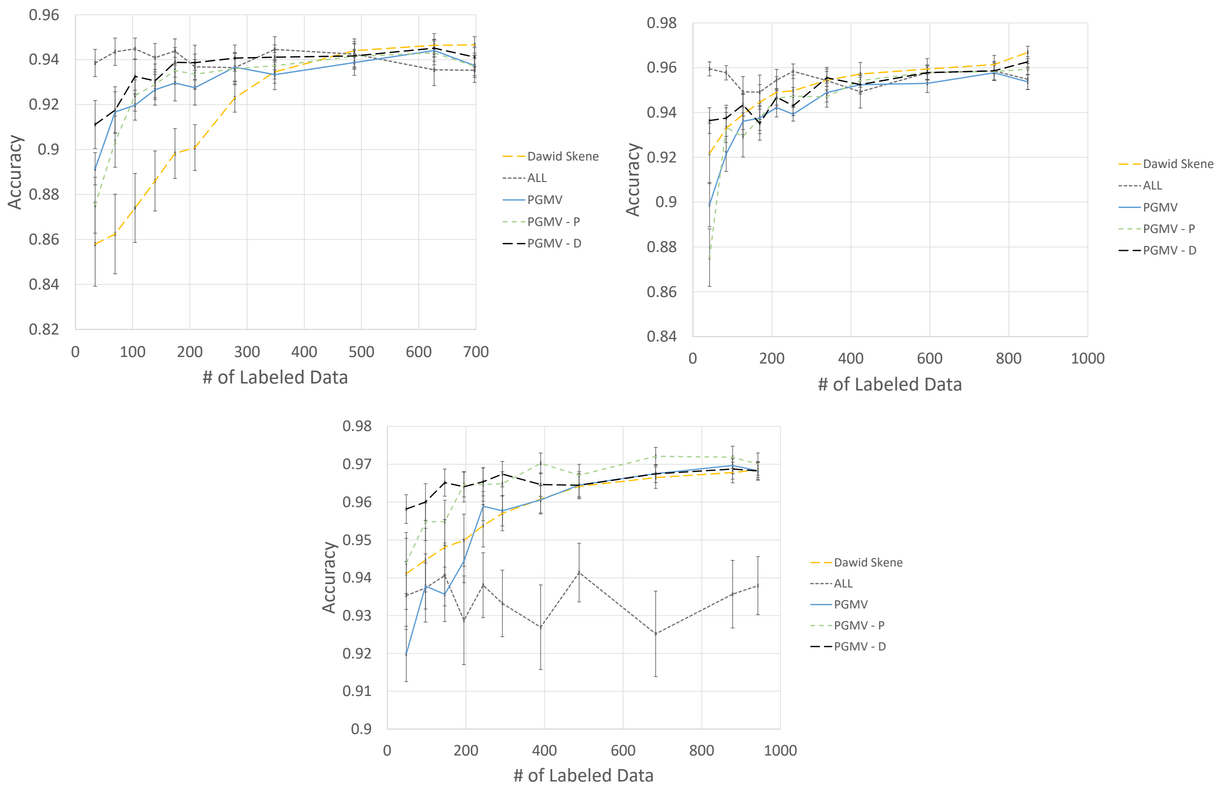

Figure 2: Additional figurs for the comparison of our algorithms and other baselines when varying the amount of labeled data. The upper left graph is averaged over the AwA2 tasks are the second group of tasks when sorted by committee potential, and the upper right graph contains the third group of tasks. The bottom graph is the fourth group of AwA2 tasks. Accuracies are reported as in the main text, computed across 3 different splits of labeled and unlabeled data, and the error bars represent the standard error. The rightmost point is the values from Table 1 and is averaged over 5 seeds.