A Proofs

In this section, we provide detailed proofs of all the results used in this manuscript. For lemma and theorem statements repeated from the main text, we add an apostrophe to indicate that it is not a new lemma/theorem being introduced.

We make use of the following result that provides an approximation bound for greedy selections for weakly submodular functions.

Lemma 10. *[Das and Kempe, 2011]) Let $\mathsf{S}^\star$ be the optimal k-sized set that maximizes $f(\cdot)$ under the k-sparsity constraint (see (2)). Let S_G be the set returned by greedy forward selection (Algorithm 1), then*

$$f(\mathsf{S}_G) \geq (1 - \exp(-\gamma_{\mathsf{S}_G,k}) f(\mathsf{S}^\star).$$

A.1 Distributed Greedy

Lemma' 2. $f(\mathsf{G}_j) \geq (1 - \exp(-\gamma_{\mathsf{G}_j,k}))f(\mathsf{T}_j)$.

Proof. From Lemma 1, we know that running greedy on $\mathsf{A}_j \cup \mathsf{T}_j$ instead of A_j will still return the set G_j,

$$\begin{aligned} f(\mathsf{G}_j) &\overset{Lemma\ 10}{\geq} (1 - \exp(-\gamma_{\mathsf{G}_j,k})) \max_{\substack{|\mathsf{S}| \leq k \\ \mathsf{S} \subset \mathsf{A}_j \cup \mathsf{T}_j}} f(\mathsf{S}) \\ &\geq (1 - \exp(-\gamma_{\mathsf{G}_j,k}))f(\mathsf{T}_j). \end{aligned}$$

□

For proving Lemma 3, we require another auxillary result.

Lemma 11. *For any $x \in \mathsf{A}, \mathbb{P}(x \in \cup_j \mathsf{G}_j) = \frac{1}{l}\sum_j \mathbb{P}(x \in \mathsf{S}_j)$.*

Proof. We have

$$\begin{aligned} &\mathbb{P}(x \in \cup_j \mathsf{G}_j) \\ &= \sum_j \mathbb{P}(x \in \mathsf{A}_i \cap x \in \textsc{Greedy}(\mathsf{A}_i, k)) \\ &= \sum_j \mathbb{P}(x \in \mathsf{A}_i)\mathbb{P}(x \in \textsc{Greedy}(\mathsf{A}_i, k) | x \in \mathsf{A}_i) \\ &= \sum_j \mathbb{P}(x \in \mathsf{A}_i)\mathbb{P}(x \in \mathsf{S}_i) \\ &= \frac{1}{l}\mathbb{P}(x \in \mathsf{S}_i). \end{aligned}$$

□

We now prove Lemma 3.

Lemma' 3. $\exists j \in [l]$, $\mathbb{E}[f(\mathsf{G})] \geq \left(1 - \frac{1}{e^{\gamma_{\mathsf{G},k}}}\right) f(\mathsf{S}_j)$.

Proof. For $i \in [k]$, let $\mathsf{B}_i : \textsc{Greedy}(\cup_j \mathsf{G}_j, i)$, so that $\mathsf{B}_k = \mathsf{G}$ in step 3 of Algorithm 2. Then,

$$\begin{aligned} &\mathbb{E}[f(\mathsf{B}_{i+1}) - f(\mathsf{B}_i)] \\ &\geq \frac{1}{k}\sum_{x \in \mathsf{A}^*} \mathbb{P}(x \in \cup_j \mathsf{G}_j)\mathbb{E}\left[f(\mathsf{B}_i \cup x) - f(\mathsf{B}_i)\right] \\ &\overset{Lemma\ 11}{=} \frac{1}{kl}\sum_{x \in \mathsf{A}^*}\left(\sum_{j=1}^{l}\mathbb{P}(x \in \mathsf{S}_i)\right)\mathbb{E}(f(\mathsf{B}_i \cup x) - f(\mathsf{B}_i)) \qquad (13) \\ &= \frac{1}{kl}\sum_{j=1}^{l}\sum_{x \in \mathsf{S}_j}\mathbb{E}(f(\mathsf{B}_i \cup x) - f(\mathsf{B}_i)) \\ &= \frac{1}{kl}\sum_{j=1}^{l}\gamma_{\mathsf{B}_i,\mathsf{S}_j \setminus \mathsf{B}_i}\mathbb{E}(f(\mathsf{B}_i \cup \mathsf{S}_j) - f(\mathsf{B}_i)) \\ &\geq \frac{1}{kl}\sum_{j=1}^{l}\gamma_{\mathsf{B}_i,\mathsf{S}_j \setminus \mathsf{B}_i}\mathbb{E}(f(\mathsf{S}_j) - f(\mathsf{B}_i)) \\ &\geq \frac{\gamma_{\mathsf{B}_i,k}}{k}\min_j \mathbb{E}(f(\mathsf{S}_j) - f(\mathsf{B}_i)). \end{aligned}$$

Using $\gamma_{\mathsf{B}_i,k} \geq \gamma_{\mathsf{G},k}$, and proceeding now as in the proof of Theorem 2, we get the desired result.

□

A.2 Stochastic Greedy

Lemma' 4. *Let $\mathsf{A}, \mathsf{B} \subset [n]$, with $|\mathsf{B}| \leq k$. Consider another set C drawn randomly from $[n] \setminus \mathsf{A}$ with $|\mathsf{C}| = \lceil \frac{n \log 1/\delta}{k} \rceil$. Then*

$$\mathbb{E}[\max_{v \in \mathsf{C}} f(v \cup \mathsf{A}) - f(\mathsf{A})] \geq \frac{(1-\delta)\gamma_{\mathsf{A},\mathsf{B} \setminus \mathsf{A}}}{k}(f(\mathsf{B}) - f(\mathsf{A})).$$

Proof. To relate the best possible marginal gain from C to the total gain of including the set $\mathsf{B} \setminus \mathsf{A}$ into A, we must upper bound the probability of overlap between C and $\mathsf{B} \setminus \mathsf{A}$ as follows:

$$\begin{aligned} \mathbb{P}(\mathsf{C} \cap (\mathsf{B} \setminus \mathsf{A}) \neq \emptyset) &= 1 - \left(1 - \frac{|\mathsf{B} \setminus \mathsf{A}|}{|[n] \setminus \mathsf{A}|}\right)^{|\mathsf{C}|} \\ &= 1 - \left(1 - \frac{|\mathsf{B} \setminus \mathsf{A}|}{|[n] \setminus \mathsf{A}|}\right)^{\lceil \frac{n \log 1/\delta}{k} \rceil} \\ &\geq 1 - \exp\left(-\frac{n \log 1/\delta}{k}\frac{|\mathsf{B} \setminus \mathsf{A}|}{|[n] \setminus \mathsf{A}|}\right) \\ &\geq 1 - \exp\left(-\frac{|\mathsf{B} \setminus \mathsf{A}| \log 1/\delta}{k}\right) \\ &\geq (1 - \exp(-\log 1/\delta))\frac{|\mathsf{B} \setminus \mathsf{A}|}{k} \qquad (14) \\ &= (1-\delta)\frac{|\mathsf{B} \setminus \mathsf{A}|}{k}, \end{aligned}$$

where (14) is because $\frac{|\mathsf{B} \setminus \mathsf{A}|}{k} \leq 1$. Let $\mathsf{S} = \mathsf{C} \cap (\mathsf{B} \setminus \mathsf{A})$. Since

$f(v \cup \mathsf{A}) - f(\mathsf{A})$ is nonnegative,

$$
\begin{aligned}
&\mathbb{E}[\max_{v \in \mathsf{C}} f(v \cup \mathsf{A}) - f(\mathsf{A})] \\
&\geq \mathbb{P}(\mathsf{S} \neq \emptyset)\mathbb{E}[\max_{v \in \mathsf{C}} f(v \cup \mathsf{A}) - f(\mathsf{A}) | \mathsf{S} \neq \emptyset] \\
&\geq (1-\delta)\frac{|\mathsf{B}\backslash\mathsf{A}|}{k}\mathbb{E}[\max_{v \in \mathsf{C}} f(v \cup \mathsf{B}) - f(\mathsf{B}) | \mathsf{S} \neq \emptyset] \\
&\geq (1-\delta)\frac{|\mathsf{B}\backslash\mathsf{A}|}{k}\mathbb{E}[\max_{v \in \mathsf{C} \cap (\mathsf{B}\backslash\mathsf{A})} f(v \cup \mathsf{A}) - f(\mathsf{A}) | \mathsf{S} \neq \emptyset] \\
&\geq (1-\delta)\frac{|\mathsf{B}\backslash\mathsf{A}|}{k}\sum_{v \in \mathsf{B}\backslash\mathsf{A}} \frac{f(v \cup \mathsf{A}) - f(\mathsf{A})}{|\mathsf{B}\backslash\mathsf{A}|} \\
&\geq \frac{(1-\delta)\gamma_{\mathsf{A},\mathsf{B}\backslash\mathsf{A}}}{k}(f(\mathsf{B}) - f(\mathsf{A})).
\end{aligned}
$$

□

A.3 Linear regression

Lemma' 6. *For the maximization of the* R^2 (7) $\nu_{\mathsf{S}} \geq \frac{\lambda_{\min}(\mathbf{C}_{\mathsf{S}})}{\lambda_{\max}(\mathbf{C}_{\mathsf{S}})}$, *where* $\mathbf{C}_{\mathsf{S}}$ *is the submatrix of* $\mathbf{C}$ *with rows and columns indexed by* S.

Proof. Say A, B is an arbitrary partition of S. Consider,

$$
\begin{aligned}
\|\mathbf{P}_{\mathsf{A}}\mathbf{y}\|_2^2 &= \mathbf{y}^\top \mathbf{P}_{\mathsf{A}}\mathbf{y} \qquad (15)\\
&= \mathbf{y}^\top \mathbf{X}_{\mathsf{A}}(\mathbf{X}_{\mathsf{A}}^\top \mathbf{X}_{\mathsf{A}})^{-1}\mathbf{X}_{\mathsf{A}}^\top \mathbf{y} \\
&\geq \|\mathbf{X}_{\mathsf{A}}^\top \mathbf{y}\|_2^2 \lambda_{\min}\left((\mathbf{X}_{\mathsf{A}}^\top \mathbf{X}_{\mathsf{A}})^{-1}\right) \\
&= \|\mathbf{X}_{\mathsf{A}}^\top \mathbf{y}\|_2^2 \frac{1}{\lambda_{\max}\left(\mathbf{X}_{\mathsf{A}}^\top \mathbf{X}_{\mathsf{A}}\right)} \\
&\geq \|\mathbf{X}_{\mathsf{A}}^\top \mathbf{y}\|_2^2 \frac{1}{\lambda_{\max}\left(\mathbf{X}_{\mathsf{S}}^\top \mathbf{X}_{\mathsf{S}}\right)}.
\end{aligned}
$$

where (15) results from the fact that all orthogonal projection matrices are symmetric and idempotent. Repeating a similar analysis for B instead of A, we get

$$
\begin{aligned}
\|\mathbf{P}_{\mathsf{A}}\mathbf{y}\|_2^2 + \|\mathbf{P}_{\mathsf{B}}\mathbf{y}\|_2^2 &\geq \frac{\|\mathbf{X}_{\mathsf{A}}^\top \mathbf{y}\|_2^2 + \|\mathbf{X}_{\mathsf{B}}^\top \mathbf{y}\|_2^2}{\lambda_{\max}\left(\mathbf{X}_{\mathsf{S}}^\top \mathbf{X}_{\mathsf{S}}\right)} \\
&= \frac{\|\mathbf{X}_{\mathsf{S}}^\top \mathbf{y}\|_2^2}{\lambda_{\max}\left(\mathbf{X}_{\mathsf{S}}^\top \mathbf{X}_{\mathsf{S}}\right)}.
\end{aligned}
$$

A similar analysis also gives $\|\mathbf{P}_{\mathsf{S}}\mathbf{y}\|_2^2 \leq \frac{\|\mathbf{X}_{\mathsf{S}}\mathbf{y}\|_2^2}{\lambda_{\min}\left(\mathbf{X}_{\mathsf{S}}^\top \mathbf{X}_{\mathsf{S}}\right)}$, which gives the desired result. □

A.4 RSC implies weak subadditivity

Let $\boldsymbol{\beta}^{(\mathsf{S})} := \max_{\text{supp}(\mathbf{x}) \in \mathsf{S}} g(\mathbf{x})$.

Lemma' 8. *For a support set* $\mathsf{S} \subset [d]$, $f(\mathsf{S}) \geq \frac{1}{2L}\|\nabla g(\mathbf{0})_{\mathsf{S}}\|_2^2$.

Proof. For any $\mathbf{v}$ with support in S,

$$
\begin{aligned}
g(\boldsymbol{\beta}^{(\mathsf{S})}) - g(\mathbf{0}) &\geq g(\mathbf{v}) - g(\mathbf{0}) \\
&\geq \langle \nabla g(\mathbf{0}), \mathbf{v} \rangle - \frac{L}{2}\|\mathbf{v}\|_2^2.
\end{aligned}
$$

Using $\mathbf{v} = \frac{1}{L}\nabla g(\mathbf{0})_{\mathsf{S}}$, we get the desired result.

□

Lemma' 9. *For any support set* S, $f(\mathsf{S}) \leq \frac{1}{2m}\|\nabla g(\mathbf{0})_{\mathsf{S}}\|_2^2$.

Proof. By strong concavity,

$$
\begin{aligned}
g(\boldsymbol{\beta}^{(\mathsf{S})}) - g(\mathbf{0}) &\leq \langle \nabla g(\mathbf{0}), \boldsymbol{\beta}^{(\mathsf{S})} \rangle - \frac{m}{2}||\boldsymbol{\beta}^{(\mathsf{S})}||_2^2 \\
&\leq \max_{\mathbf{v}} \langle \nabla g(\mathbf{0}), \mathbf{v}) \rangle - \frac{m}{2}\|\mathbf{v}\|_F^2,
\end{aligned}
$$

where $\mathbf{v}$ is an arbitrary vector that has support only on S. Optimizing the RHS over $\mathbf{v}$ gives the desired result. □

B Additional experiments

Figures 4,5 shows the performance of all algorithms on the following metrics: log likelihood (normalized with respect to a null model), generalization to new test measurements from the same true support parameter, area under ROC, and percentage of the true support recovered for $l = 2$. Recall that Figure 1 presents the results from the same experiment with $l = 10$. Clearly, the greedy algorithms benefit more from increased number of partitions.

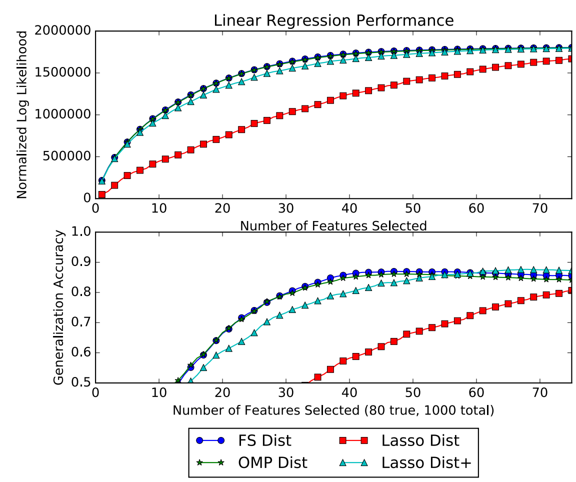

Figure 4: Distributed linear regression, $l = 2$ partitions, $n = 800$ training and test samples, $\alpha = 0.5$. Training/testing performance

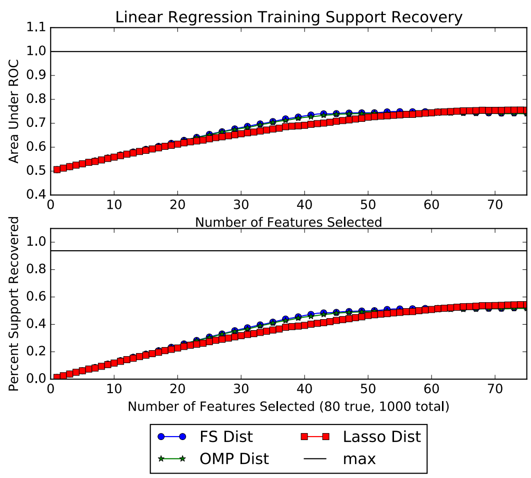

Figure 5: Distributed linear regression, $l = 2$ partitions, $n = 800$ training and test samples, $\alpha = 0.5$. Support Recovery Performance