PRELIMINARY VERSION: DO NOT CITE
The AAAI Digital Library will contain the published
version some time after the conference

Model-sharing Games: Analyzing Federated Learning Under Voluntary Participation

Kate Donahue,[1] **Jon Kleinberg,** [1, 2]
[1] Department of Computer Science, Cornell University
[2] Department of Information Science, Cornell University
kdonahue@cs.cornell.edu, kleinber@cs.cornell.edu

Abstract

Federated learning is a setting where agents, each with access to their own data source, combine models learned from local data to create a global model. If agents are drawing their data from different distributions, though, federated learning might produce a biased global model that is not optimal for each agent. This means that agents face a fundamental question: should they join the global model or stay with their local model? In this work, we show how this situation can be naturally analyzed through the framework of coalitional game theory.

Motivated by these considerations, we propose the following game: there are heterogeneous players with different model parameters governing their data distribution and different amounts of data they have noisily drawn from their own distribution. Each player's goal is to obtain a model with minimal expected mean squared error (MSE) on their own distribution. They have a choice of fitting a model based solely on their own data, or combining their learned parameters with those of some subset of the other players. Combining models reduces the variance component of their error through access to more data, but increases the bias because of the heterogeneity of distributions. In this work, we derive exact expected MSE values for problems in linear regression and mean estimation. We use these values to analyze the resulting game in the framework of hedonic game theory; we study how players might divide into coalitions, where each set of players within a coalition jointly construct model(s). We analyze three methods of federation, modeling differing degrees of customization. In uniform federation, the agents collectively produce a single model. In coarse-grained federation, each agent can weight the global model together with their local model. In fine-grained federation, each agent can flexibly combine models from each other agent in the federation. For each method, we constructively analyze the stable partitions of players into coalitions.

1 Introduction

Imagine a situation as follows: a hospital is trying to evaluate the effectiveness of a certain procedure based on data it has collected from procedures done on patients in their facilities. It seems likely that certain attributes of the patient influences the effectiveness of the procedure, so the hospital analysts opt to fit a linear regression model with parameters $\hat{\boldsymbol{\theta}}$. However, because of the limited amount of data the hospital has access to, this model has relatively high error. Luckily, other hospitals also have data from implementations of this same procedure. However, for reasons of privacy, data incompatibility, data size, or other operational considerations, the hospitals don't wish to share raw patient data. Instead, they they opt to combine their models by taking a weighted average of the parameters learned by each hospital. If there are M hospitals and hospital i has n_i samples, the combined model parameters would look like:

$$\hat{\boldsymbol{\theta}}^f = \frac{1}{\sum_{i=1}^{M} n_i} \sum_{i=1}^{M} \hat{\boldsymbol{\theta}}_i \cdot n_i$$

The situation described above could be viewed as a stylized model of *federated learning*. Federated learning is a distributed learning process that is currently experiencing rapid innovations and widespread implementation (Li et al. 2020; Kairouz et al. 2019). It is used in cases where data is distributed across multiple agents and cannot be combined centrally for training. For example, federated learning is implemented in word prediction on cell phones, where transferring the raw text data would be infeasible given its large size (and sensitive content). The motivating factor for using federated learning is that access to more data will reduce the variance in a learned model, reducing its error.

However, there could be a downside to using federated learning. In the hospital example, it seems quite reasonable that certain hospitals might have different true generating models for their data, based on the differences in patient populations or variants of the procedure implementation, for example. Two dissimilar hospitals that are federating together will see a decrease in their model's error due to model variance, but an increase in their error due to model bias. This raises some fundamental questions for each participating hospital - or, more generally, each agent i considering federated learning. Which other agents should i federate with in order to minimize its error? Will those other agents be interested in federating with i? Does there exist some stable arrangement of agents into federating clusters, and if so, what does that arrangement look like?

Numerous works have explored the issue of heterogeneous data in federated learning - we discuss specifically how they relate to ours in a later section. Often the goal in these lines of work is to achieve equality in error rates guaranteed to each agent, potentially by actively collecting

more data or using transfer learning to ensure the model better fits local data. However, to our knowledge, there has not yet been work that systematically looks at the participation questions inherent in federated learning through the lens of game theory — especially the theory of *hedonic games*, which studies the formation of self-sustaining coalitions.

In a hedonic game, players are grouped together into clusters or coalitions: the overall collection of coalitions is called a *coalition structure*. Each player's utility depends solely on the identity of the other players in its coalition. A common question in hedonic games is the stability of a coalition structure. A coalition structure Π is *core-stable* (or "in the core") if there does not exist a coalition C so that every player in C prefers C to its coalition in Π. A coalition structure is *strictly core stable* if there does not exist a coalition C so that every player in C weakly prefers C to its coalition in Π, and at least one player $\in C$ strictly prefers C to Π. A coalition structure is *individually stable* if there does not exist a coalition $C \in \Pi$ so that a player $i \notin C$ prefers $C \cup \{i\}$ to its arrangement in Π and all players in C weakly prefer $C \cup \{i\}$ to C (Bogomolnaia and Jackson 2002).

To explain the analogy of federated learning to hedonic games, we first consider that each agent in federated learning is a player in a hedonic game. A player is in coalition with other players if it is federating with them. Its cost is its expected error in a given federating cluster, which depends only on the identity of other players in its federating cluster. Players are assumed to be able to move between federating clusters only if doing so would benefit itself and not harm other players in the cluster it is moving to: notably, we allow players to freely leave a cluster, even if doing so would harm the players in the cluster it leaves behind.

The present work: Analyzing federated learning through hedonic game theory[1] In this work, we use the framework of hedonic games to analyze the stability of coalitions in data-sharing applications that capture key issues in federated learning. By working through a sequence of deliberately stylized models, we obtain some general insights about participation and stability in these kinds of applications.

For the first case, we analyze *uniform* federation. In this simplest case, a federating cluster produces a single model, which each player uses. For uniform federation, first we consider the case where all players have the same number of data points. We show that in this game, when the number of data points n is fairly small, the only core-stable coalition structure is to have all players federating together, in the "grand coalition". When n is large, the only core-stable coalition structure is to have all players separate (doing local learning). There exists a point case of intermediate n size where all coalition structures are core-stable. Next, we analyze the case where all players have either one of two sizes ("small" or "large"). The analysis is more complicated, but we demonstrate constructively that there always exists an individually stable partition of players into clusters.

[1]For the full paper, please refer to (Donahue and Kleinberg 2020).

Besides uniform federation, we also analyze two other forms of federation. For *coarse-grained* federation, the federating cluster still produces a single model, but each player can weight the global model with their local model, allowing some personalization. For coarse-grained federation, when all players have the same number of samples, we show that the grand coalition (all players federating together) is always the only core stable arrangement. For the small/large case, we produce a simple, efficient algorithm to calculate a strictly core-stable arrangement. Additionally, we show that, for this federation method, the grand coalition is always individually stable (no player wishes to unilaterally defect).

Finally, for *fine-grained* federation, each player is allowed to take the local models of other players in the federation and combine them using whichever weights they choose to produce a model customized for their use. With fine-grained federation, we show that the grand coalition is always core stable.

We are only able to produce these hedonic game theory results because of our derivations of exact error values for the underlying inference problems. We calculate these values for all three methods of federation, and for agents federating in two situation: 1) a mean estimation problem and 2) a linear regression problem. The error values depend on the number of samples each agent has access to, with the expectation taken over the values of samples each agent draws as well as the possible different true parameters of the data each player is trying to model. Our results are completely independent of the generating distributions used, relying only weakly on two parameters.

The results in this paper are theoretical and do not depend on any experiments or code. However, while writing the paper, we found it useful to write and work with code to check conjectures. Some of that code is publicly available at `github.com/kpdonahue/model_sharing_games`.

Before moving to the main technical content, the next section will walk through a motivating example, followed by a review of related literature and a description of our model and assumptions. Beyond technical assumptions, recent work (Cooper (2020)) has highlighted the importance of describing *normative* assumptions researchers make: we also include a summary of the most important normative assumptions of our analysis in our ethics statement after the main text.

Related works

Incentives and federated learning: Blum et al. (2017) describes an approach to handling heterogeneous data where more samples are iteratively gathered from each agent in a way so that all agents are incentivized to participate in the grand coalition during federated learning. Duan et al. (2021) builds a framework to schedule data augmentation and resampling. Yu, Bagdasaryan, and Shmatikov (2020) demonstrates empirically that there can be cases where individuals get lower error with local training than federated and evaluates empirical solutions. Wang et al. (2020) analyzes the question of when it makes sense to split or not to split datasets drawn from different distributions. Finally, Blum

Coalition structure	$err_a(\cdot)$	$err_b(\cdot)$	$err_c(\cdot)$
$\{a\},\{b\},\{c\}$	2	2	2
$\{a,b\},\{c\}$	1.5	1.5	2
$\{a,b,c\}$	1.3	1.3	1.3

Table 1: The expected errors using uniform federation of players in each coalition when all three players have 5 samples each, with parameters $\mu_e = 10, \sigma^2 = 1$. Each row denotes a different coalition partition: for example, $\{a,b\}\{c\}$ indicates that players a and b are federating together while c is alone.

Coalition structure	$err_a(\cdot)$	$err_b(\cdot)$	$err_c(\cdot)$
$\{a\},\{b\},\{c\}$	2	2	0.4
$\{a,b\},\{c\}$	1.5	1.5	0.4
$\{a\},\{b,c\}$	2	1.72	0.39
$\{a,b,c\}$	1.55	1.55	0.41

Table 2: The expected errors using uniform federation of players in each coalition when players a and b have 5 samples each and player c has 25 samples, with parameters $\mu_e = 10, \sigma^2 = 1$.

et al. (2020) analyzes notions of envy and efficiency with respect to sampling allocations in federated learning.

Transfer learning: Mansour et al. (2020) and Deng, Kamani, and Mahdavi (2020) both propose theoretical methods for using transfer learning to minimize error provided to agents with heterogeneous data. Li et al. (2019) and Martinez, Bertran, and Sapiro (2020) both provide methods to produce a more uniform level of error rates across agents participating in federated learning.

Clustering and federated learning: Sattler, Muller, and Samek (2020) and Shlezinger, Rini, and Eldar (2020) provide an algorithm to "cluster" together players with similar data distributions with the aim of providing them with lower error. They differ from our approach in that they consider the case where there is some knowledge of each player's data distribution, where we only assume knowledge of the number of data points. Additionally, their approach doesn't explicitly consider agents to be game-theoretic actors in the same way that this one does. Interestingly, Guazzone, Anglano, and Sereno (2014) uses a game theoretic framework to analyze federated learning, but with the aim of minimizing energy usage, not error rate.

2 Motivating example

To motivate our problem and clarify the types of analyses we will be exploring, we will first work through a simple mean estimation example. (The Github repository contains numerical calculations for this section.) Calculating the error each player can expect requires two parameters: μ_e, which reflects the average error each player experiences when sampling data from its own personal distribution, and σ^2, which reflects the average variance in the true parameters between players. In this section we will use $\mu_e = 10, \sigma^2 = 1$, but will discuss later how to handle when they may be imperfectly known.

Coalition structure	$err_a(\cdot)$	$err_b(\cdot)$	$err_c(\cdot)$
$\{a\},\{b\},\{c\}$	0.4	0.4	0.4
$\{a,b\},\{c\}$	0.7	0.7	0.4
$\{a,b,c\}$	0.8	0.8	0.8

Table 3: The expected errors using uniform federation of players in each coalition when players a, b, c each have 25 samples, with parameters $\mu_e = 10, \sigma^2 = 1$.

Coalition structure	$err_a(\cdot)$	$err_d(\cdot)$
$\{a\},\{b\},\{c\},\{d\}$	0.333	0.0333
$\{a,b,c\},\{d\}$	0.278	0.0333
$\{a,b,c,d\}$	0.280	0.0326

Table 4: The expected errors using optimal coarse-grained federation when players a, b, c each have 30 samples, while player d has 300 samples, with parameters $\mu_e = 10, \sigma^2 = 1$.

First, we will analyze uniform federation, with three players, a, b, and c. We will first assume that each player has 5 samples from their local data distribution: Table 1 gives the error each player can expect in this situation. Arrangements equivalent up to renaming of players are omitted. Every player sees its error minimized in the "grand coalition" π_g where all three players are federating together. This implies that the only arrangement that is stable (core-stable or individually stable) is π_g.

Next, we assume that player c increases the amount of samples it has from 5 to 25: Table 2 demonstrates the error each player can expect in this situation. Here, the players have different preferences over which arrangement they would most prefer. The "small" players a and b would most prefer $\{a,b\}\{c\}$, whereas the "large" player c would most prefer $\{a\},\{b,c\}$ or (identically) $\{b\},\{a,c\}$. However out of all of these coalition structures, only $\{a,b\},\{c\}$ is stable (either core stable or individually stable). Note that $\{a\},\{b,c\}$ is not stable because the coalition $C = \{a,b\}$ is one where each player prefers C to its current situation.

Thirdly, we will assume that all three players have 25 samples: this example is shown in Table 3. As in Table 1, the players have identical preferences. However, in this case, the players minimize their error by being alone. Overall, stability results from this example are part of a broader pattern we will analyze in later sections.

Next, we will explore the two other methods of federation: coarse-grained and fine-grained. Both offer some degree of personalization, with varying levels of flexibility.

Table 4 shows an example using coarse federation with four players: three have 30 samples each, and the fourth player has 300 samples. We assume the weights w_j are set optimally for each player. For conciseness, some columns and rows are omitted. Note that both player types get lower error in π_g than they would with local learning: that is, π_g is individually stable (stable against defections of any player alone). However, it is also clear that π_g is not core stable: in particular, the three small players would get lower error in $\{a,b,c\}$ than in π_g. These results will be examined theoretically in later sections: with optimal weighting, coarse-grained federation will always have an individually stable π_g that is not necessarily core stable.

Coalition structure	$err_a(\cdot)$	$err_d(\cdot)$
$\{a\},\{b\},\{c\},\{d\}$	0.333	0.0333
$\{a,b,c\},\{d\}$	0.278	0.0333
$\{a,b,c,d\}$	0.269	0.0325

Table 5: The expected errors using optimal fine-grained federation when players a, b, c each have 30 samples, while player d has 300 samples, with parameters $\mu_e = 10, \sigma^2 = 1$.

Finally, we examine fine-grained federation. Table 5 analyzes the same case as coarse-grained federation previously, but with optimally-weighted fine-grained federation. The full error table shows that π_g *is* core stable because each player minimizes their error in this arrangement. This result will hold theoretically: when optimal fine-grained federation is used, π_g always minimizes error for every player and is thus core stable.

In later sections we will give theoretical results that explain this example more fully, but understanding the core-stable partitions here will help to build intuition for more general results.

3 Model and assumptions

This section introduces our model. We assume that there is a fixed set of $[M]$ players, and player j has a fixed number of samples, n_j. Though the number of samples is fixed, it is possible to analyze a varying number of samples by investigating all games involving the relevant number of samples. Each player draws their true parameters i.i.d. (independent and identically distributed) $(\theta_j, \epsilon_j^2) \sim \Theta$. ϵ_j^2 represents the amount of noise in the sampling process for a given player.

In the case of mean estimation, θ_j is a scalar representing the true mean of player j. Player j draws samples i.i.d. from its true distribution: $Y \sim \mathcal{D}_j(\theta_j, \epsilon_j^2)$. Samples are drawn with variance ϵ_j^2 around the true mean of the distribution.

In the case of linear regression, $\boldsymbol{\theta}_j$ is a D-dimensional vector representing the coefficients on the true classification function, which is also assumed to be linear. Each player draws n_j input datapoints from their own input distribution $\boldsymbol{X}_j \sim \mathcal{X}_j$ such that $\mathbb{E}_{x\sim\mathcal{X}_j}[\boldsymbol{x}^T\boldsymbol{x}] = \Sigma_j$. They then noisily observe the outputs, drawing values i.i.d. $\boldsymbol{Y}_j \sim \mathcal{D}_j(\boldsymbol{X}_j^T\boldsymbol{\theta}_j, \epsilon_j^2)$, where ϵ_j^2 again denotes the variance of how samples are drawn around the true mean.

There are three methods of federation. In uniform federation, a single model is produced for all members of the federating coalition:

$$\hat{\boldsymbol{\theta}}^f = \frac{1}{\sum_{i=1}^M n_i} \sum_{i=1}^M \hat{\boldsymbol{\theta}}_i \cdot n_i$$

In coarse-grained federation, each player has a parameter w_j that it uses to weight the global model with its own local model, producing an averaged model:

$$\hat{\theta}_j^w = w_j \cdot \hat{\theta}_j + (1 - w_j) \cdot \frac{1}{N} \sum_{i=1}^M \hat{\theta}_i \cdot n_i$$

for $w_j \in [0, 1]$. Note that $w_j = 0$ corresponds to unweighted federated learning and $w_j = 1$ corresponds to pure local learning. Finally, with fine-grained federation, each player j as a vector of weights $\boldsymbol{v}_j$ that they use to weight every other player's contribution to their estimate:

$$\hat{\theta}_j^v = \sum_{i=1}^M v_{ji}\theta_i$$

for $\sum_{i=1}^M v_{ji} = 1$. Note that we can recover the w weighting case with $v_{jj} = w + \frac{(1-w)\cdot n_j}{N}$ and $v_{ji} = (1 - w) \cdot \frac{n_i}{N}$. Coarse-grained and fine-grained federation each have player-specific parameters (w, v) that can be tuned. When those parameters are set optimally for the given player, we refer to the models as "optimal" coarse-grained or fine-grained federation. We will prove in later sections how to calculate optimal weights.

We denote $\mu_e = \mathbb{E}_{(\theta_i,\epsilon_i^2)\sim\Theta}[\epsilon_i^2]$: the expectation of the error parameter. In the mean estimation case, $\sigma^2 = Var(\theta_i)$ represents the variance around the mean. In the linear regression case, $\sigma_d^2 = Var(\boldsymbol{\theta}_i^d)$ for $d \in [D]$.

We assume that each player knows how many samples it has access to. It may or may not have access to the data itself, but it does not know how its values (or its parameters) differ from the mean. For example, it does not know if the data it has is unusually noisy or if its true mean lies far from the true mean of other players.

All of the stability analysis results depend on the parameters μ_e and σ^2. However, the reliance is fairly weak: often the player only needs to whether the number of samples they have n_j is larger or smaller than the ratio $\frac{\mu_e}{\sigma^2}$.

Much of this paper analyzes the stability of coalition structures. Analyzing stability could be relevant because players can actually move between coalitions. However, even if players aren't able to actually move, analyzing the stability of a coalition tells us something about its optimality for each set of players.

4 Expected error results

This paper's first contribution is to derive exact expected values for the MSE of players under different situations. The fact that these values are exact allows us to precisely reason about each player's incentives in later sections. We will state the theorems here and provide the proofs in the full version (Donahue and Kleinberg 2020).

The approach for this section was first to derive expected MSE values for the most general case and then derive values for other cases as corollaries. The most general case is linear regression with fine-trained federation. First, we note that we can derive coarse-grained or uniform federation by setting the v_{ji} weights to the appropriate values. Next, we note that mean estimation is a special case of linear regression. For intuition, consider a model where a player draws an x value that is deterministically 1, then multiplies it by an unknown single parameter θ_j, then then takes a measurement y of this mean with noise ϵ_j^2. This corresponds exactly to the mean estimation case, where a player has a true mean θ_j and observes y as a sample, with noise ϵ_j^2. We can use this representation to simplify the error terms, with more details given in the full version (Donahue and Kleinberg 2020).

First, we give the expected MSE for local estimation:

Theorem 4.1. *For linear regression, the expected MSE of local estimation for a player with n_j samples is*

$$\mu_e \cdot tr\left[\Sigma_j \mathbb{E}_{X_j \sim \mathcal{X}_j}\left[\left(\boldsymbol{X}_j^T \boldsymbol{X}_j\right)^{-1}\right]\right]$$

If the distribution of input values $\mathcal{X}_j$ is a D-dimensional multivariate normal distribution with 0 mean, then, the expected MSE of local estimation can be simplified to:

$$\frac{\mu_e}{n_j - D - 1} D$$

In the case of mean estimation, the error term can be simplified to:

$$\frac{\mu_e}{n_j}$$

Next, we calculate the expected MSE for fine-grained federation:

Theorem 4.2. *For linear regression with fine-grained federation, the expected MSE of federated estimation for a player with n_j samples is:*

$$L_j + \left(\sum_{i \neq j} v_{ji}^2 + \left(\sum_{i \neq j} v_{ji}\right)^2\right) \cdot \sum_{d=1}^{D} \mathbb{E}_{x \sim \mathcal{X}_j}[(\boldsymbol{x}^d)^2] \cdot \sigma_d^2$$

where L_j is equal to:

$$\mu_e \sum_{i=1}^{M} v_{ji}^2 \cdot tr[\Sigma_j \mathbb{E}_{Y \sim \mathcal{D}(\theta_i, \epsilon_i^2)}\left[(\boldsymbol{X}_i^T \boldsymbol{X}_i)^{-1}\right]$$

If the distribution of input values $\mathcal{X}_i$ is a D-dimensional multivariate normal distribution with 0 mean, this can be simplified to:

$$\mu_e \sum_{i=1}^{M} v_{ji}^2 \cdot \frac{D}{n_i - D - 1}$$

In the case of mean estimation, the entire error term can be simplified to:

$$\mu_e \sum_{i=1}^{M} v_{ji}^2 \cdot \frac{1}{n_i} + \left(\sum_{i \neq j} v_{ji}^2 + \left(\sum_{i \neq j} v_{ji}\right)^2\right) \cdot \sigma^2$$

Finally, we derive as corollaries the expected MSE for the uniform federation and the coarse-grained case.

Corollary 4.3. *For uniform linear regression, the expected MSE of federated estimation for a player with n_j samples is:*

$$L_j + \frac{\sum_{i \neq j} n_i^2 + (N - n_j)^2}{N^2} \sum_{d=1}^{D} \mathbb{E}_{x \sim \mathcal{X}_j}[(\boldsymbol{x}^d)^2] \cdot \sigma_d^2$$

where L_j is equal to:

$$\mu_e \sum_{i=1}^{M} \frac{n_i^2}{N^2} tr[\Sigma_j \mathbb{E}_{Y \sim \mathcal{D}(\theta_i, \epsilon_i^2)}\left[(\boldsymbol{X}_i^T \boldsymbol{X}_i)^{-1}\right]$$

or, if the distribution of input values $\mathcal{X}_i$ is a D-dimensional multivariate normal distribution with 0 mean, can be simplified to

$$\mu_e \sum_{i=1}^{M} \frac{n_i^2}{N^2} \frac{D}{n_i - D - 1}$$

In the case of mean estimation, the entire error term can be simplified to:

$$\frac{\mu_e}{N} + \frac{\sum_{i \neq j} n_i^2 + (N - n_j)^2}{N^2} \sigma^2$$

where $N = \sum_{i=1}^{M} n_i$.

Corollary 4.4. *For coarse-grained linear regression, the expected MSE of federated estimation for a player with n_j samples is:*

$$L_j + (1-w)^2 \cdot \frac{\sum_{i \neq j} n_i^2 + (N - n_j)^2}{N^2} \sum_{d=1}^{D} \mathbb{E}_{x \sim \mathcal{X}_j}[(\boldsymbol{x}^d)^2] \cdot \sigma_d^2$$

where L_j is equal to:

$$\mu_e \cdot (1-w)^2 \cdot \sum_{i=1}^{M} \frac{n_i^2}{N^2} tr[\Sigma_j \mathbb{E}_{Y \sim \mathcal{D}(\theta_i, \epsilon_i^2)}\left[(\boldsymbol{X}_i^T \boldsymbol{X}_i)^{-1}\right] +$$

$$\mu_e \left(w^2 + 2\frac{(1-w)w \cdot n_j}{N}\right) \cdot tr[\Sigma_j \mathbb{E}_{Y \sim \mathcal{D}(\theta_i, \epsilon_i^2)}\left[(\boldsymbol{X}_j^T \boldsymbol{X}_j)^{-1}\right]$$

or, if the distribution of input values $\mathcal{X}_i$ is a D-dimensional multivariate normal distribution with 0 mean, can be simplified to

$$\mu_e \cdot (1-w)^2 \cdot \sum_{i=1}^{M} \frac{n_i^2}{N^2} \frac{D}{n_i - D - 1}$$
$$+\mu_e \cdot \left(w^2 + 2 \cdot \frac{(1-w) \cdot w \cdot n_j}{N}\right) \cdot \frac{D}{n_i - D - 1}$$

In the case of mean estimation, the entire error term can be simplified to:

$$\mu_e \left(\frac{w^2}{n_j} + \frac{1-w^2}{N}\right) + \frac{\sum_{i \neq j} n_i^2 + (N - n_j)^2}{N^2} \cdot (1-w)^2 \sigma^2$$

where $N = \sum_{i=1}^{M} n_i$.

The exact MSE for linear regression follows a very similar form to that for mean estimation. In all cases, the bias component (the term involving σ_d^2) is in the exact same form and could be directly modified to mean estimation by using $\sigma^{2\prime} = \sum_{d=1}^{D} \mathbb{E}_{x \sim \mathcal{X}_j}[(\boldsymbol{x}^d)^2] \cdot \sigma_d^2$. The variance component (the term involving μ_e) fits the exact form of mean estimation in the limit where $n_j >> D$. In this case, the error can be modified to fit mean estimation by using $\mu_e' = D \cdot \mu_e$. This approximation is good when there are many more samples than the dimension of the linear regression problem under investigation: for most cases of model fitting, this assumption is reasonable.

For the rest of the paper, we will use the $n_j >> D$ assumption: consequentially, *all of our results apply equally to linear regression and mean estimation*.

5 Uniform federation: coalition formation

In this section, we analyze the stability of coalition structures in the case that uniform federation is used. We consider two cases: 1) where all players have the same number of datapoints n and 2) where all players have either a "small" or "large" number of points. We will use π_l to refer to the coalition partition where all players are alone and π_g to refer to the grand coalition. Proofs from this section are given in the full version (Donahue and Kleinberg 2020).

All players have the same number of samples

In this case, the analysis simplifies greatly:

Lemma 5.1. *If all players have the same number of samples n, then:*

- *If $n < \frac{\mu_e}{\sigma^2}$, players minimize their error in π_g.*
- *If $n > \frac{\mu_e}{\sigma^2}$, players minimize their error in π_l.*
- *If $n = \frac{\mu_e}{\sigma^2}$, players are indifferent between any arrangement of players.*

Proof. In the case that all players have the same number of samples, we can use $n_i = n$ to simplify the error term:

$$\frac{\mu_e}{M \cdot n} + \sigma^2 \frac{M-1}{M}$$

In order to see whether players would prefer a larger group (higher M) or a smaller group (smaller M), we take the derivative of the error with respect to M:

$$-\frac{\mu_e}{M^2 \cdot n} + \frac{\sigma^2}{M^2} = \frac{\sigma^2 \cdot n - \mu_e}{n \cdot M^2}$$

This is positive when $n > \frac{\mu_e}{n}$: a player gets higher error the more players it is federating with. This is negative when $n < \frac{\mu_e}{\sigma^2}$: a player gets lower error the more players it is federating with. This is 0 when $n = \frac{\mu_e}{\sigma^2}$, which implies players should be indifferent between different arrangements. Plugging in for $n = \frac{\mu_e}{\sigma^2}$ in the error equation gives $\frac{\mu_e \cdot \sigma^2}{M \cdot \mu_e} + \sigma^2 \frac{M-1}{M} = \sigma^2$ which is equivalent to the error a player would get alone: $\frac{\mu_e}{n} = \frac{\mu_e \cdot \sigma^2}{\mu_e} = \sigma^2$. □

As a corollary, we can classify the core stable arrangements cleanly:

Corollary 5.2. *For uniform federation, if all players have the same number of samples n, then:*

- *If $n < \frac{\mu_e}{\sigma^2}$, π_g is the only partition that is core-stable.*
- *If $n > \frac{\mu_e}{\sigma^2}$, π_l is the only partition that is core-stable.*
- *If $n = \frac{\mu_e}{\sigma^2}$, any arrangement of players is core-stable.*

Small & large player case

In this section, we add another layer of depth by allowing players to come in one of two "sizes". "Small" players have n_s samples and "large" ones have n_ℓ samples, with $n_s < n_\ell$. We demonstrate that versions of the game in this pattern always have a stable partition by constructively producing an element that is stable. Note that this is *not* true in general of hedonic games. As discussed in Bogomolnaia and Jackson (2002), there are multiple instances where a game might have no stable partition.

To characterize this space, we divide it into cases depending on the relative size of n_s, n_ℓ. We will use the notation $\pi(s, \ell)$ to denote a coalition with s small players and ℓ large players, out of a total of S and L present. We will use $\pi(s_1, \ell_1) \succ_S \pi(s_2, \ell_2)$ to mean that the small players prefer coalition $\pi(s_1, \ell_1)$ to $\pi(s_2, \ell_2)$ and $\pi(s_1, \ell_1) \succ_L \pi(s_2, \ell_2)$ to mean the same preference, but for large players.

Case 1: $n_s, n_\ell \geq \frac{\mu_e}{\sigma^2}$ The first case is when n_s is large: it turns out that each player minimizes their error by using local learning, which means that π_l is in the core. The lemma below is more general than the small/large case, but implies that when $n_s > \frac{\mu_e}{\sigma^2}$, π_l is the only element in the core and when $n_s = \frac{\mu_e}{\sigma^2}$ then any arrangement where the large players are alone are is in the core.

Lemma 5.3. *For uniform federation, if $n_i > \frac{\mu_e}{\sigma^2}$ for all $i \in [M]$, then π_l is the unique element in the core.*
If $n_i \geq \frac{\mu_e}{\sigma^2}$ for all $i \in [M]$, with $n_k > \frac{\mu_e}{\sigma^2}$ for at least one player k, then any arrangement where the players with samples $n_k > \frac{\mu_e}{\sigma^2}$ are alone is in the core.

Case 2: $n_s, n_\ell \leq \frac{\mu_e}{\sigma^2}$ Next, we consider the case where both the small and large players have a relatively small number of samples. In this situation, it turns out that the grand coalition is core stable.

Theorem 5.4. *For uniform federation, if $n_\ell \leq \frac{\mu_e}{\sigma^2}$ and $n_s < n_\ell$, then the grand coalition π_g is core stable.*

Case 3: $n_s < \frac{\mu_e}{\sigma^2}$**,** $n_\ell > \frac{\mu_e}{\sigma^2}$ Finally, we consider the case where the small players have a number of samples below the $\frac{\mu_e}{\sigma^2}$ boundary, while the large players have a number of samples above this threshold.

Theorem 5.5. *Assume uniform federation with $n_\ell > \frac{\mu_e}{\sigma^2}$. Then, there exists an arrangement of small and large players that is individually stable and a computationally efficient algorithm to calculate it.*

The proof of Theorem 5.5 is constructive: it gives an exact arrangement that is individually stable. One natural question is whether this arrangement is also core stable. The answer to this question is "no": we show that this arrangement can fail to be core stable. This avenue is explained more in the full version (Donahue and Kleinberg 2020).

6 Coarse-grained federation

In this section, we analyze coarse-grained federation. As a reminder, in this situation, each player has a parameter w_j that it uses to weight the global model with its own local model.

$$\hat{\theta}_j^w = w_j \cdot \hat{\theta}_j + (1 - w_j) \cdot \frac{1}{N} \sum_{i=1}^{M} \hat{\theta}_i \cdot n_i$$

for $w_j \in [0, 1]$. All proofs from this section are given in the full version (Donahue and Kleinberg 2020).

Note that the w_j value is a parameter that each player can set independently. The lemma below analyzes the optimal value of w_j and tells us that each player would prefer federation, in some form, to being alone.

Lemma 6.1. *For coarse-grained federation, the minimum error is always achieved when $w_j < 1$, implying that federation is always preferable to local learning.*

Corollary 6.2. *For coarse-grained federation, when w_j is set optimally, the grand coalition π_g is always individually stable.*

Specifically, this means that no player wishes to unilaterally deviate from π_g. However, this does *not* mean that each player prefers the grand coalition π_g to some other federating coalition. For example, refer to Section 2 for an example where the grand coalition π_g is not core stable.

In the rest of this section, we will analyze the stability of coalition structures in the that the w parameters are set optimally (optimal coarse-grained federation). First, we will find it useful to get the closed-form value for expected MSE of a player using optimal coarse-grained federation:

Lemma 6.3. *A player using coarse-grained federation parameter has expected MSE:*

$$\frac{\mu_e \cdot (N - n_j) + (\sum_{i \neq j} n_i^2 + (N - n_j)^2) \cdot \sigma^2}{(N - n_j) \cdot N + n_j \cdot (\sum_{i \neq j} n_i^2 + (N - n_j)^2) \cdot \frac{\sigma^2}{\mu_e}}$$

where $N = \sum_{i=1}^{M} n_i$.

All players have the same number of samples

Lemma 6.4 is the analog to Lemma 5.1 in the previous section. Here, the results differ: with optimal coarse-grained federation, the grand coalition π_g is *always* the only stable arrangement, no matter how small or large n is relative to $\frac{\mu_e}{\sigma^2}$.

Lemma 6.4. *For mean estimation with coarse-grained federation, if $n_j = n$, then π_g is the only element in the core.*

Proof. Using the error term derived in Lemma 6.3, plugging in for $n_i = n$ and simplifying gives:

$$\frac{\frac{\mu_e^2}{n \cdot M} + \mu_e \cdot \sigma^2}{\mu_e + n \cdot \sigma^2}$$

As M increases, the error (numerator) decreases always - so π_g is where each player minimizes their error and is thus core stable. □

Small & large player case

In this subsection, we similarly extend results for the "small" and "large" case that was introduced in the previous section. The analysis turns out to be much simpler than in the uniform federation case, and also produce stronger results: strict core stability, rather than individual stability.

Theorem 6.5. *If optimal coarse-grained federation is used, then:*

- *If $\pi_g \preceq_S \pi(S, 0)$ (small player weakly prefers $\pi(S, 0)$), then $\{\pi(S, 0), \pi(0, L)\}$ is strictly core stable.*
- *If $\pi_g \succ_S \pi(S, 0)$ (small player strictly prefers π_g), then π_g is strictly core stable.*

7 Fine-grained federation

In this section, we analyze fine-grained federation. As a reminder, with this method, each player j as a vector of weights $\boldsymbol{v}_j$ that they use to weight every other player's contribution to their estimate.

$$\hat{\theta}_j^v = \sum_{i=1}^{M} v_{ji} \theta_i$$

for $\sum_{i=1}^{M} v_{ji} = 1$.

We calculate the optimal $\boldsymbol{v}$ weights for player j's error.

Lemma 7.1. *Define $V_i = \sigma^2 + \frac{\mu_e}{n_i}$. Then, the value of $\{v_{ji}\}$ that minimizes player j's error is:*

$$v_{jj} = \frac{1 + \sigma^2 \sum_{i \neq j} \frac{1}{V_i}}{1 + V_j \sum_{i \neq j} \frac{1}{V_i}}$$

$$v_{jk} = \frac{1}{V_k} \cdot \frac{V_j - \sigma^2}{1 + V_j \sum_{i \neq j} \frac{1}{V_i}} \quad k \neq j$$

The proof of this lemma is given in the full version (Donahue and Kleinberg 2020).

From this analysis, a few properties become clear. To start with, v_{jj} and v_{jk} are always strictly between 0 and 1. This implies the following lemma:

Corollary 7.2. *With optimal fine-grained federation, π_g is optimal for each player.*

Proof. Suppose by contradiction that some other coalition π' gave player j a lower error. WLOG, assume this coalition omitted player k. In this case, the v weights for π' can be represented as a length M vector with 0 in the kth entry. However, set of weights is achievable in π_g: it is always an option to set a player's coefficient v_{jk} equal to 0. This contradicts the use of $\boldsymbol{v}_j$ as an optimal weighting, so it cannot be the case that any player gets lower error in a different coalition. □

Similarly, the fact that π_g is optimal for every player implies that it is in the core, and that it is the only element in the core.

8 Conclusions and future directions

In this work, we have drawn a connection between a simple model of federated learning and the game theoretic tool of hedonic games. We used this tool to examine stable partitions of the space for two variants of the game. In service of this analysis, we computed exact error values for mean estimation and linear regression, as well as for three different variations of federation.

We believe that this framework is a simple and useful tool for analyzing the incentives of multiple self-interested agents in a learning environment. There are many fascinating extensions. For example, completely characterizing the core (including whether it is always non-empty) in the case of arbitrary number of samples $\{n_i\}$ is an obvious area of investigation. Besides this, it could be interesting to compute exact or approximate error values for cases beyond mean estimation and linear regression.

Normative assumptions (ethics statement) This paper is primarily descriptive: it aims to model a phenomenon in the world, not to say whether that phenomenon is good or bad. For example, it could be that society as a whole values situations where many players federate together and might wish to require players to do so, regardless of whether this minimizes their error. It might be the case that society prefers all players, regardless of how many samples they have access to, have roughly similar error rates. Our use of the expected mean squared error is also worth reflecting on: it assumes that over- and under-estimates are equally costly and that larger mis-estimates are more costly. In a more subtle point, we are taking the expected MSE over parameter draws $\mathbb{E}_{(\theta_i,\epsilon_i^2)\sim\Theta}$. A player with a true mean that happens to fall far from the mean might experience a much higher error than its expected MSE.

In the entirely of this paper, we are taking as fixed the requirement that data not be shared, either for privacy or technical capability reasons, and so are implicitly valuing that requirement more than the desire for lower error. We are also assuming that the problem at hand is completely encompassed by the machine learning task, which might omit the fact that non-machine learning solutions may be better suited. It also may be the fact that technical requirements other than error rate are more important: for example, the desire to balance the amount of computation done by each agent.

Acknowledgments

This work was supported in part by a Simons Investigator Award, a Vannevar Bush Faculty Fellowship, a MURI grant, AFOSR grant FA9550-19-1-0183, grants from the ARO and the MacArthur Foundation, and NSF grant DGE-1650441. We are grateful to A. F. Cooper, Thodoris Lykouris, Hakim Weathersppon, and the AI in Policy and Practice working group at Cornell for invaluable discussions. In particular, we thank A.F. Cooper for discussions around normative assumptions. Finally, we are grateful to Katy Blumer for discussions around code in the Github repository.

References

Abu-Mostafa, Y.; Lin, H.; and Magdon-Ismail, M. 2012. Learning from data: a short course: AMLbook. *Google Scholar* .

Anderson, T. W. 1962. An introduction to multivariate statistical analysis. Technical report, Wiley New York.

Blum, A.; Haghtalab, N.; Procaccia, A. D.; and Qiao, M. 2017. Collaborative PAC Learning. In Guyon, I.; Luxburg, U. V.; Bengio, S.; Wallach, H.; Fergus, R.; Vishwanathan, S.; and Garnett, R., eds., *Advances in Neural Information Processing Systems 30*, 2392–2401. Curran Associates, Inc. URL http://papers.nips.cc/paper/6833-collaborative-pac-learning.pdf.

Blum, A.; Haghtalab, N.; Shao, H.; and Phillips, R. L. 2020. Unpublished work, private correspondence.

Bogomolnaia, A.; and Jackson, M. O. 2002. The Stability of Hedonic Coalition Structures. *Games and Economic Behavior* 38(2): 201 – 230. ISSN 0899-8256. doi: https://doi.org/10.1006/game.2001.0877. URL http://www.sciencedirect.com/science/article/pii/S0899825601908772.

Casella, G. 1992. Illustrating empirical Bayes methods. *Chemometrics and intelligent laboratory systems* 16(2): 107–125.

Cooper, A. F. 2020. Where Is the Normative Proof? Assumptions and Contradictions in ML Fairness Research.

Deng, Y.; Kamani, M. M.; and Mahdavi, M. 2020. Adaptive Personalized Federated Learning.

Donahue, K.; and Kleinberg, J. 2020. Model-sharing Games: Analyzing Federated Learning Under Voluntary Participation.

Duan, M.; Liu, D.; Chen, X.; Liu, R.; Tan, Y.; and Liang, L. 2021. Self-Balancing Federated Learning With Global Imbalanced Data in Mobile Systems. *IEEE Transactions on Parallel and Distributed Systems* 32(1): 59–71.

Efron, B.; and Morris, C. 1977. Stein's paradox in statistics. *Scientific American* 236(5): 119–127.

Guazzone, M.; Anglano, C.; and Sereno, M. 2014. A Game-Theoretic Approach to Coalition Formation in Green Cloud Federations. *2014 14th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing* doi: 10.1109/ccgrid.2014.37. URL http://dx.doi.org/10.1109/CCGrid.2014.37.

Kairouz, P.; McMahan, H. B.; Avent, B.; Bellet, A.; Bennis, M.; Bhagoji, A. N.; Bonawitz, K.; Charles, Z.; Cormode, G.; Cummings, R.; D'Oliveira, R. G. L.; Rouayheb, S. E.; Evans, D.; Gardner, J.; Garrett, Z.; Gascón, A.; Ghazi, B.; Gibbons, P. B.; Gruteser, M.; Harchaoui, Z.; He, C.; He, L.; Huo, Z.; Hutchinson, B.; Hsu, J.; Jaggi, M.; Javidi, T.; Joshi, G.; Khodak, M.; Konečný, J.; Korolova, A.; Koushanfar, F.; Koyejo, S.; Lepoint, T.; Liu, Y.; Mittal, P.; Mohri, M.; Nock, R.; Özgür, A.; Pagh, R.; Raykova, M.; Qi, H.; Ramage, D.; Raskar, R.; Song, D.; Song, W.; Stich, S. U.; Sun, Z.; Suresh, A. T.; Tramèr, F.; Vepakomma, P.; Wang, J.; Xiong, L.; Xu, Z.; Yang, Q.; Yu, F. X.; Yu, H.; and Zhao, S. 2019. Advances and Open Problems in Federated Learning.

Li, T.; Sahu, A. K.; Talwalkar, A.; and Smith, V. 2020. Federated Learning: Challenges, Methods, and Future Directions. *IEEE Signal Processing Magazine* 37(3): 50–60. ISSN 1558-0792. doi:10.1109/msp.2020.2975749. URL http://dx.doi.org/10.1109/MSP.2020.2975749.

Li, T.; Sanjabi, M.; Beirami, A.; and Smith, V. 2019. Fair Resource Allocation in Federated Learning.

Mansour, Y.; Mohri, M.; Ro, J.; and Suresh, A. T. 2020. Three Approaches for Personalization with Applications to Federated Learning.

Martinez, N.; Bertran, M.; and Sapiro, G. 2020. Minimax Pareto Fairness: A Multi Objective Perspective. In *Proceedings of the 37th International Conference on Machine Learning*, Proceedings of Machine Learning Research. PMLR. URL https://proceedings.icml.cc/static/paper_files/icml/2020/1084-Paper.pdf.

Morris, C. N. 1986. *Empirical Bayes: a frequency-Bayes compromise*, volume Volume 8 of *Lecture Notes–Monograph Series*, 195–203. Hayward, CA: Institute of Mathematical Statistics. doi:10.1214/lnms/1215540299. URL https://doi.org/10.1214/lnms/1215540299.

Paquay, P. 2018. Learning-from-data-Solutions. URL https://github.com/ppaquay/Learning-from-Data-Solutions.

Sattler, F.; Muller, K.-R.; and Samek, W. 2020. Clustered Federated Learning: Model-Agnostic Distributed Multitask Optimization Under Privacy Constraints. *IEEE Transactions on Neural Networks and Learning Systems* 1–13. ISSN 2162-2388. doi:10.1109/tnnls.2020.3015958. URL http://dx.doi.org/10.1109/TNNLS.2020.3015958.

Sellentin, E.; and Heavens, A. F. 2015. Parameter inference with estimated covariance matrices. *Monthly Notices of the Royal Astronomical Society: Letters* 456(1): L132–L136. ISSN 1745-3925. doi:10.1093/mnrasl/slv190. URL https://doi.org/10.1093/mnrasl/slv190.

Shlezinger, N.; Rini, S.; and Eldar, Y. C. 2020. The Communication-Aware Clustered Federated Learning Problem. In *2020 IEEE International Symposium on Information Theory (ISIT)*, 2610–2615.

Wang, H.; Hsu, H.; Diaz, M.; and Calmon, F. P. 2020. To Split or Not to Split: The Impact of Disparate Treatment in Classification.

Yu, T.; Bagdasaryan, E.; and Shmatikov, V. 2020. Salvaging Federated Learning by Local Adaptation.