

Federated Curvature & Multi-Task Federated Machine Learning

Théo LELEU

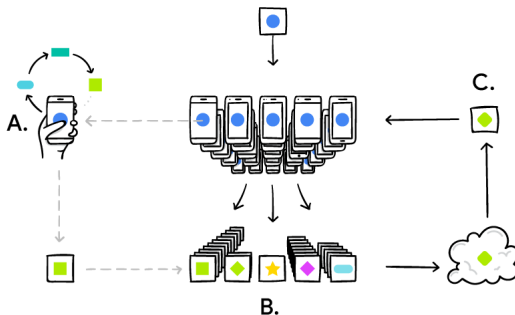
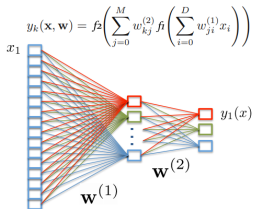
Christophe Cerisara
LORIA
Mines Nancy

11 Juin 2020



Réseaux de Neurones et Federated learning

« Multi-layer perceptron » (MLP)



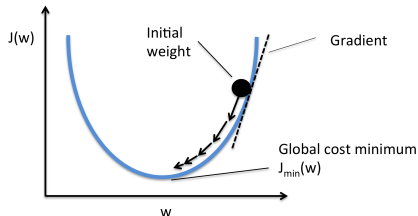
Descente de Gradient Stochastique

Soit un réseau avec des poids W et une vitesse d'apprentissage ϵ .

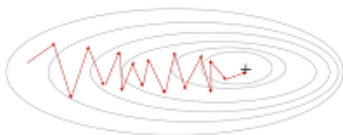
$$W = W - \epsilon \times \nabla L(D_i) \text{ (E itérations)}$$

avec L la fonction de perte, souvent :

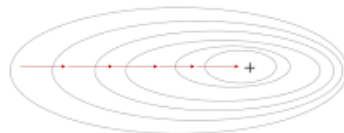
$$L(D_i) = \frac{1}{N} \sum_{D_i} (\text{Observation} - \text{Prédiction})^2$$



Stochastic Gradient Descent

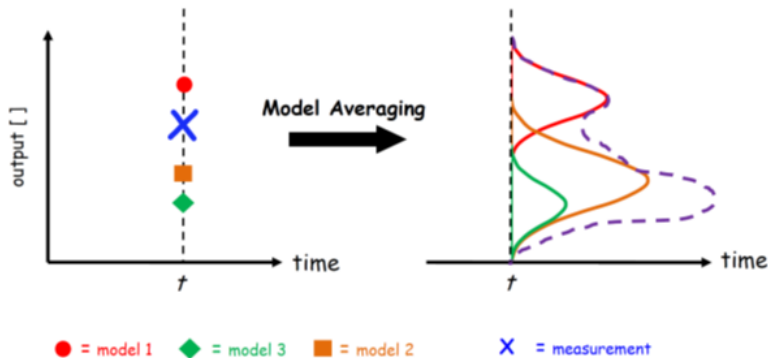


Gradient Descent



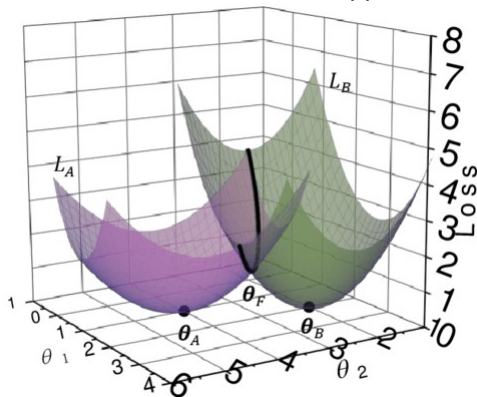
Federated Averaging

Pour gérer la fusion, nous moyennons les modèles des clients pour obtenir le modèle général.



Federated Curvature

Sur un modèle à plusieurs clients,
on aura des résultats sur les clients distincts de l'optimum mais, on
cherche à s'en rapprocher.



On utilise une méthode de pénalité pour forcer cela :

$$\tilde{L}_B(w) \approx L_B(w) + \lambda(w - \hat{w}_A)^T \text{diag}(H_{L_A})(w - \hat{w}_A) \text{ with } H_{L_A} = \frac{\partial^2 L_A}{\partial x_i \partial x_j}$$

D'où pour un modèle à n clients à l'étape t :

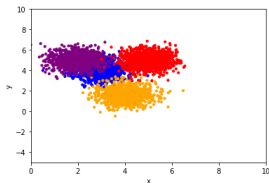
En entrée sur s :

- $L_{t,s}(w) = \hat{L}_{t,s}(w) + \lambda w^T \times (u_t w - 2v_t)$
- Chaque appareil s reçoit
 $u_t = \sum_{j \in S \setminus s} \text{diag}(H_{t-1,j})$ et $v_t = \sum_{j \in S \setminus s} \text{diag}(H_{t-1,j}) w_{t-1,j}$

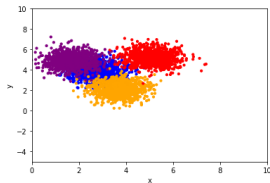
En sortie :

- Il faut transmettre $w_{t,j}$ et $\text{diag}(H_{L_{t,j}})$

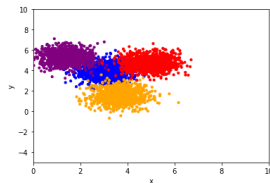
Expériences



Client 1

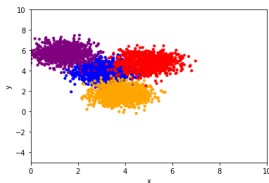


Client 2



Client 3

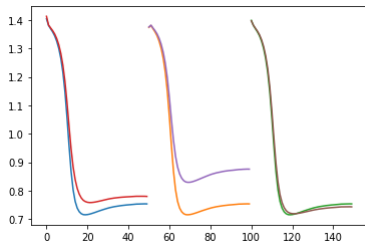
- Nous étudions trois groupes de points représentant des corpus d'apprentissage et observons l'impact du passage de FedAVG à FedCurv.



Tâche de test avec les quatres catégories de points

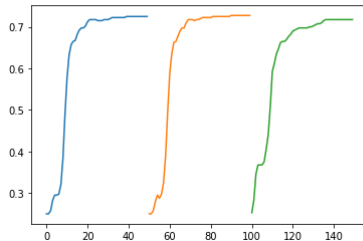
FedAVG

Loss



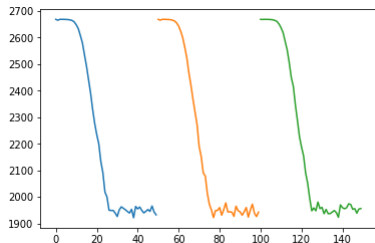
Accuracy

acc finale machine 0 0.725
acc finale machine 1 0.7275
acc finale machine 2 0.7175



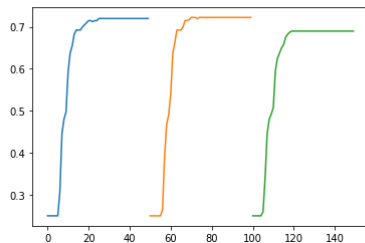
FedCurv

Loss



Accuracy

acc finale machine 0 0.72
acc finale machine 1 0.7225
acc finale machine 2 0.69



Etude de tâches de sentiments : Senteval

name	N	task	C	examples	label(s)
MR	11k	sentiment (movies)	2	"Too slow for a younger crowd , too shallow for an older one."	neg
CR	4k	product reviews	2	"We tried it out christmas night and it worked great ."	pos
SUBJ	10k	subjectivity/objectivity	2	"A movie that doesn't aim too high , but doesn't need to."	subj
MPQA	11k	opinion polarity	2	"don't want"; "would like to tell";	neg, pos
TREC	6k	question-type	6	"What are the twin cities ?"	LOC:city
SST-2	70k	sentiment (movies)	2	"Audrey Tautou has a knack for picking roles that magnify her [...]"	pos
SST-5	12k	sentiment (movies)	5	"nothing about this movie works."	0

Table 1: **Classification tasks.** C is the number of classes and N is the number of samples.


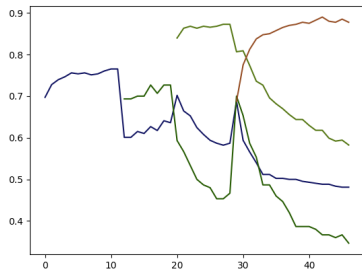
name	N	task	output	premise	hypothesis	label
SNLI	560k	NLI	3	"A small girl wearing a pink jacket is riding on a carousel."	"The carousel is moving."	entailment
SICK-E	10k	NLI	3	"A man is sitting on a chair and rubbing his eyes"	"There is no man sitting on a chair and rubbing his eyes"	contradiction
SICK-R	10k	STS	[0, 5]	"A man is singing a song and playing the guitar"	"A man is opening a package that contains headphones"	1.6
STS14	4.5k	STS	[0, 5]	"Liquid ammonia leak kills 15 in Shanghai"	"Liquid ammonia leak kills at least 15 in Shanghai"	4.6
MRPC	5.7k	PD	2	"The procedure is generally performed in the second or third trimester."	"The technique is used during the second and, occasionally, third trimester of pregnancy."	paraphrase
COCO	565k	ICR	sim		"A group of people on some horses riding through the beach."	rank

Table 2: **Natural Language Inference and Semantic Similarity tasks.** NLI labels are contradiction, neutral and entailment. STS labels are scores between 0 and 5. PD=paraphrase detection, ICR=image-caption retrieval.

Figure: Description des données présentes dans le package Senteval de Conneau et Kiela

Etude de tâches de sentiments : Senteval

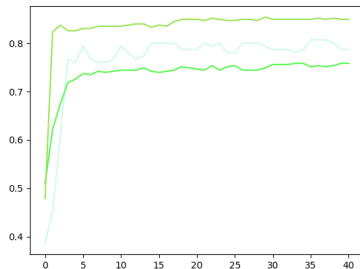
SGD



Accuracy de SGD

Par ordre : CR, MR, MPQA, SUBJ

FedAVG



Accuracy de FedAVG

De bas en haut MR, CR, MPQA

Merci pour votre attention !