

TP7

On détaille et met en oeuvre divers algorithmes de descente.

- Charger `~despres/ML3_data.py`, et construire un échantillonnage excessivement serré (4000 samples) de la fonction $f^{\text{obj}}(x) = x - x^7$.
- Charger `~despres/ML3.py` dont les paramètres sont réglés pour que l'optimiseur soit l'algorithme de gradient (avec initialisation aléatoire et LReLU).

On commencera par analyser la documentation dont une copie est au lien `~despres/SGD.pdf`

```
-----  
w = w - learning_rate * g  
-----  
velocity = momentum * velocity - learning_rate * g  
w = w + velocity  
-----  
velocity = momentum * velocity - learning_rate * g  
w = w + momentum * velocity - learning_rate * g  
-----
```

Lancez le soft `~despres/ML3.py`. Qu'observe-t-on?

- Utilisez les 3 algorithmes, méthode de gradient, méthode de gradient avec moment, puis avec l'accélération de Nesterov. Les choses s'améliorent-elles vraiment?
- Utilisez l'optimiseur Adam et vérifiez que cela résout les problèmes de convergence.

Ajouter des couches et essayer d'augmenter la précision finale.