

Introduction:

Exploration et application des techniques d'optimisation en apprentissage automatique : *descente de gradient en lots (BGD)* et *descente de gradient stochastique (SGD)*. Étude détaillée de ces méthodes, leurs différences et leur application pratique. Ajustement d'un nuage de points à une hyperquadrique et classification des données sur le cancer du sein. Objectif : améliorer nos compétences en apprentissage automatique et contribuer à la recherche.

Objectifs

- Comparer les performances de BGD et SGD dans différents scénarios.
- Appliquer BGD et SGD à des problèmes réels (ajustement de courbes et classification de données sur le cancer du sein).
- Mettre en évidence l'importance du choix de la méthode d'optimisation adaptée.

BGD vs SGD

$$\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

BGD

for $i : 1, \dots, m$

$$\theta_j := \theta_j - \alpha (h_{\theta}(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

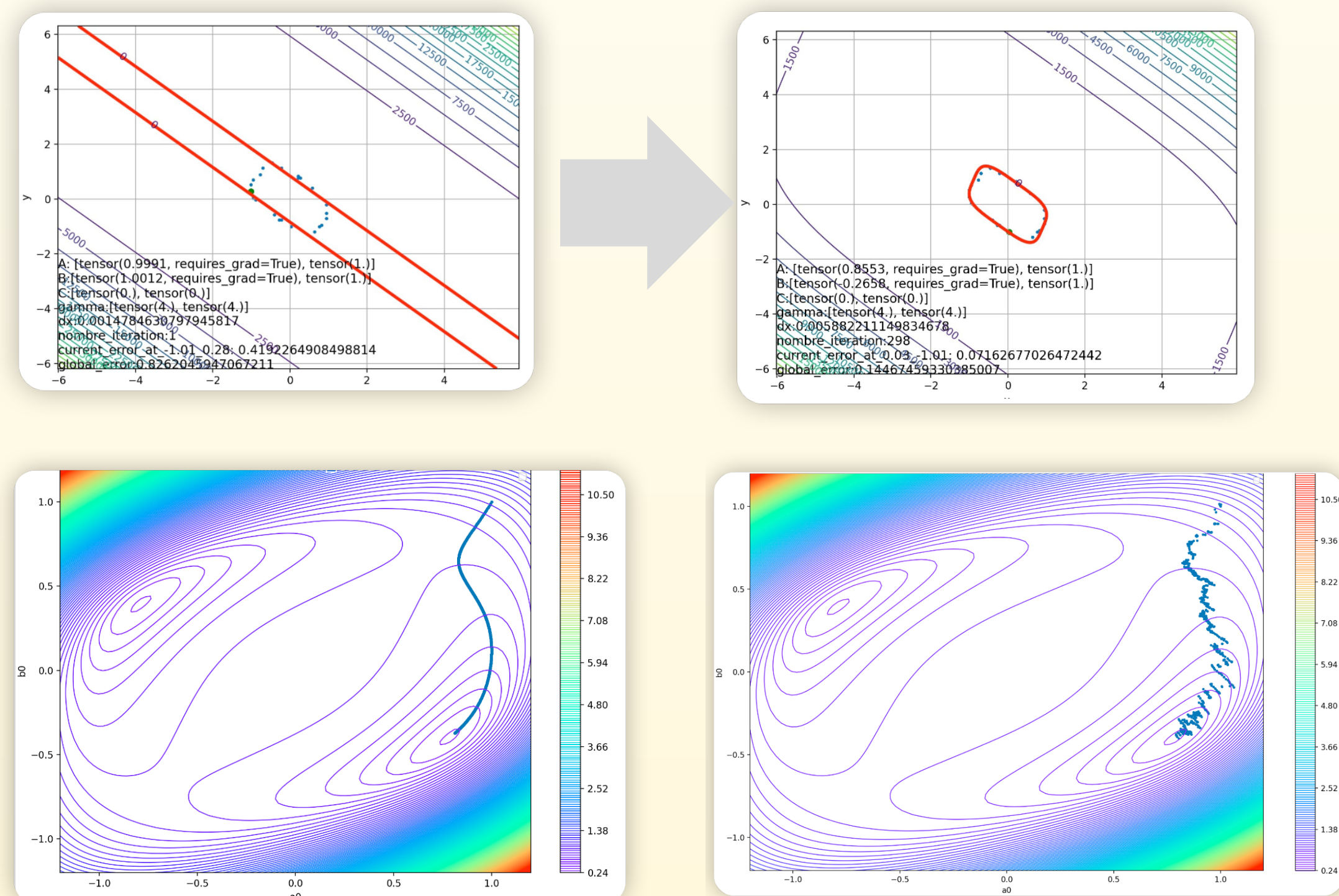
(for every $j = 0, \dots, m$)

SGD

SGD et BGD sont deux méthodes couramment utilisées en apprentissage automatique pour optimiser les modèles. SGD met à jour les poids en utilisant un seul exemple à la fois, ce qui est rapide mais introduit de la variabilité. BGD traite tous les exemples à la fois, offrant une mise à jour plus stable mais peut être plus lente avec de nombreux exemples.

Résultats

1. fitter un nuage de points par une hyperquadrique

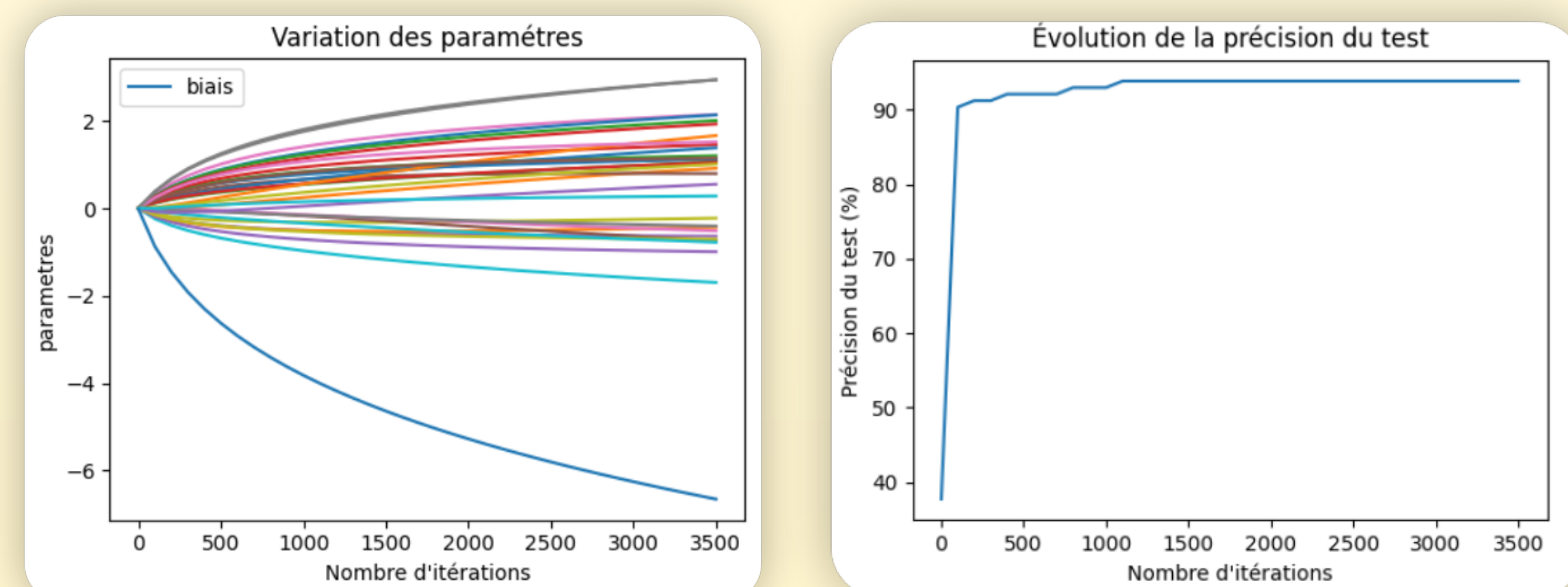


BGD

vs

SGD

2. classification des données relatives au cancer du sein



Conclusion

- La SGD présente de meilleures performances en termes de traitement de grands ensembles de données et de convergence rapide, tandis que la BGD a l'avantage de la stabilité et de la précision.
- Le choix d'un point de départ et d'un taux d'apprentissage appropriés est crucial.
- Il est important d'ajuster de manière adaptative le taux d'apprentissage pour améliorer la vitesse de convergence.