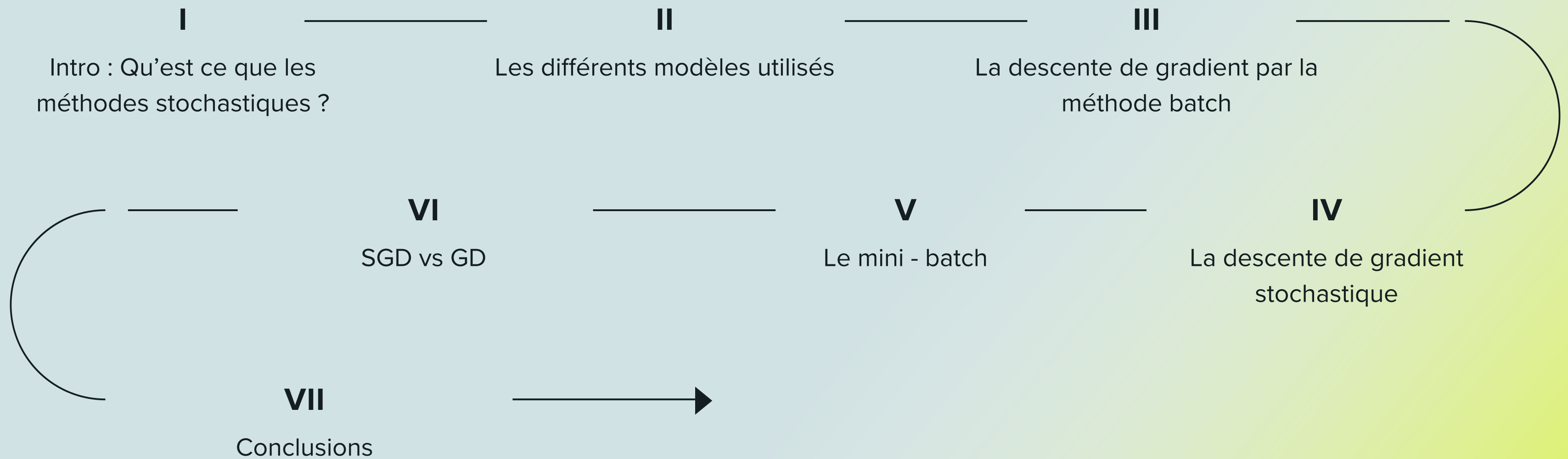


# Les méthodes stochastiques

# Sommaire



I

**Introduction : Qu'est ce que les méthodes stochastiques ?**

# Présentation

Le terme "stochastique" se réfère à l'introduction de l'incertitude et du hasard dans le processus d'apprentissage. Dans le contexte de l'apprentissage automatique, les méthodes stochastiques sont des approches d'optimisation qui diffèrent des méthodes déterministes traditionnelles. Elles introduisent une dose d'aléatoire dans le choix des instances de données traitées à chaque itération, permettant ainsi d'atteindre des solutions plus rapidement et avec une efficacité accrue, notamment sur de vastes ensembles de données.

- **Une Approche dynamique et adaptable.**
- **Un Rôle crucial dans la recherche de solutions optimales**
- **Une évolution significative dans la résolution de problèmes complexes d'optimisation**

# II

## Les différents modèles utilisés

# Les modèles stochastiques

## La descente de gradient stochastique

Utilisé pour l'optimisation de fonctions, notamment dans les modèles linéaires comme SGDRegressor et SGDClassifier.

## Perceptron Stochastique

Un modèle linéaire pour la classification binaire qui utilise la descente de gradient stochastique.

## Réseaux de Neurones Stochastiques

Les réseaux de neurones avec des éléments stochastiques, tels que les dropouts pendant l'entraînement, sont considérés comme stochastiques.

## Machines à Vecteurs de Support Stochastiques (SGD-SVM)

Une variante stochastique des machines à vecteurs de support pour la classification.

# Les modèles stochastiques

## **K-Means Stochastique**

Une variante stochastique de l'algorithme K-Means pour le clustering.

## **Descente de Gradient Bayésienne Stochastique (SGD-Bayesien)**

Une approche bayésienne stochastique pour l'optimisation.

## **Inferential Stochastic Ensemble (ISE)**

Un modèle stochastique pour l'inférence bayésienne.

## **Forêt Stochastique (Stochastic Forest)**

Une version stochastique des forêts aléatoires.



# III

## La descente de gradient par la méthode batch



# Principe de fonctionnement

- Les gradients de la fonction coût sont calculés en utilisant l'ensemble des données d'entraînement, à chaque étape de la Descente de Gradient par batch
- Ces gradients représentent la direction dans laquelle les paramètres du modèle doivent être ajustés pour minimiser la fonction de coût.

# Avantages et limites

## *Avantages*

### **Convergence assurée**

Elle converge généralement vers le minimum global de la fonction de coût, garantissant ainsi une solution précise.

### **Utilisation efficace des ressources**

L'ensemble des données est exploité, avantageux pour dataframes de petites et moyennes tailles

## *Limites*

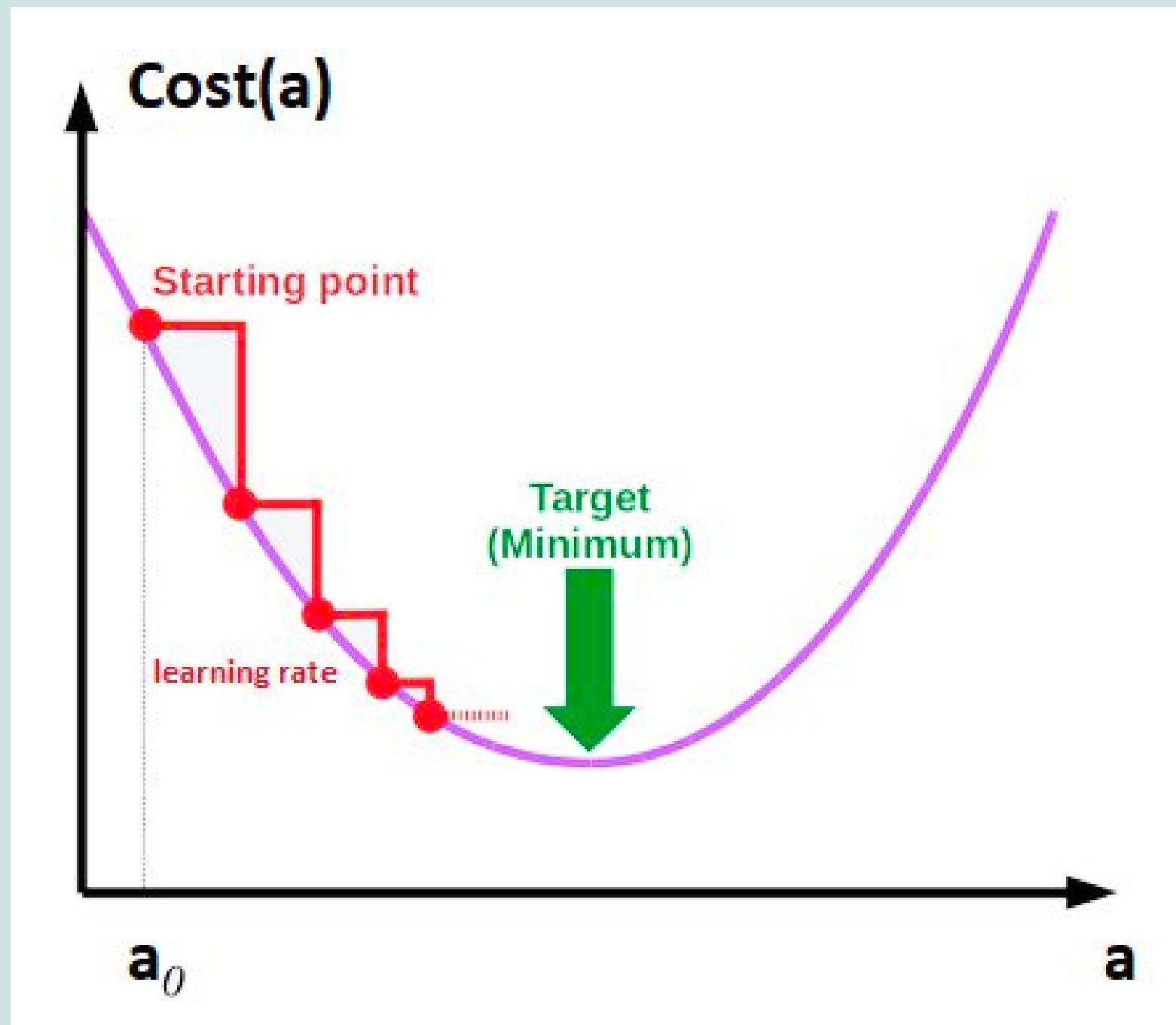
### **Calcul intensif**

Prohibitif pour jeu de données massif

### **Sensibilité aux outliers**

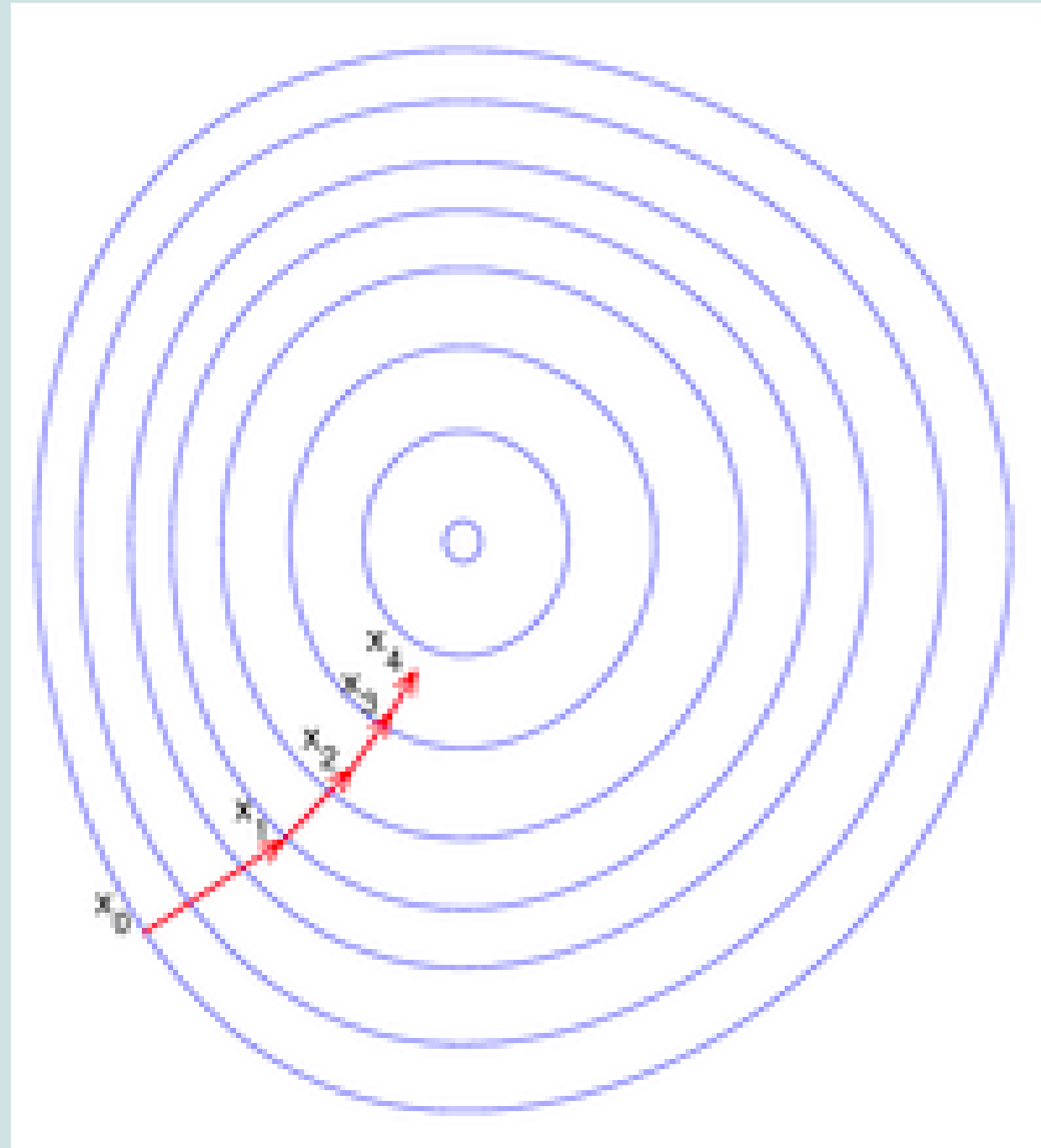
Impact significatif la mise à jour des paramètres, peut affecter la convergence

# Visualisation



Dans un espace paramétrique, la Descente de Gradient par batch suit une trajectoire plus lisse et directe vers le minimum global, illustrant sa convergence déterministe.

# Visualisation



# Choix du learning rate

La sélection d'un learning rate approprié est fondamentale pour la convergence.

Un taux trop élevé peut entraîner une oscillation autour du minimum, tandis qu'un taux trop faible peut ralentir la convergence.

# IV

## La descente de gradient stochastique (SGD)

# Principe de fonctionnement

- Contrairement à la Descente de Gradient par Batch, la SGD sélectionne aléatoirement une seule instance de données à chaque itération pour calculer les gradients et ajuster les paramètres du modèle.
- Cette approche rend le processus plus rapide, car il n'est pas nécessaire de parcourir l'ensemble complet des données.



# Avantages et limites

## *Avantages*

### **Efficacité sur des Ensembles Massifs**

En traitant une seule instance à la fois, la SGD est performante sur des ensembles de données massifs où le coût computationnel serait prohibitif pour la Descente de Gradient par batch

### **Adaptabilité aux Changements Rapides**

La SGD peut s'adapter rapidement à des changements dans les données, rendant cette méthode utile dans des scénarios dynamiques.

## *Limites*

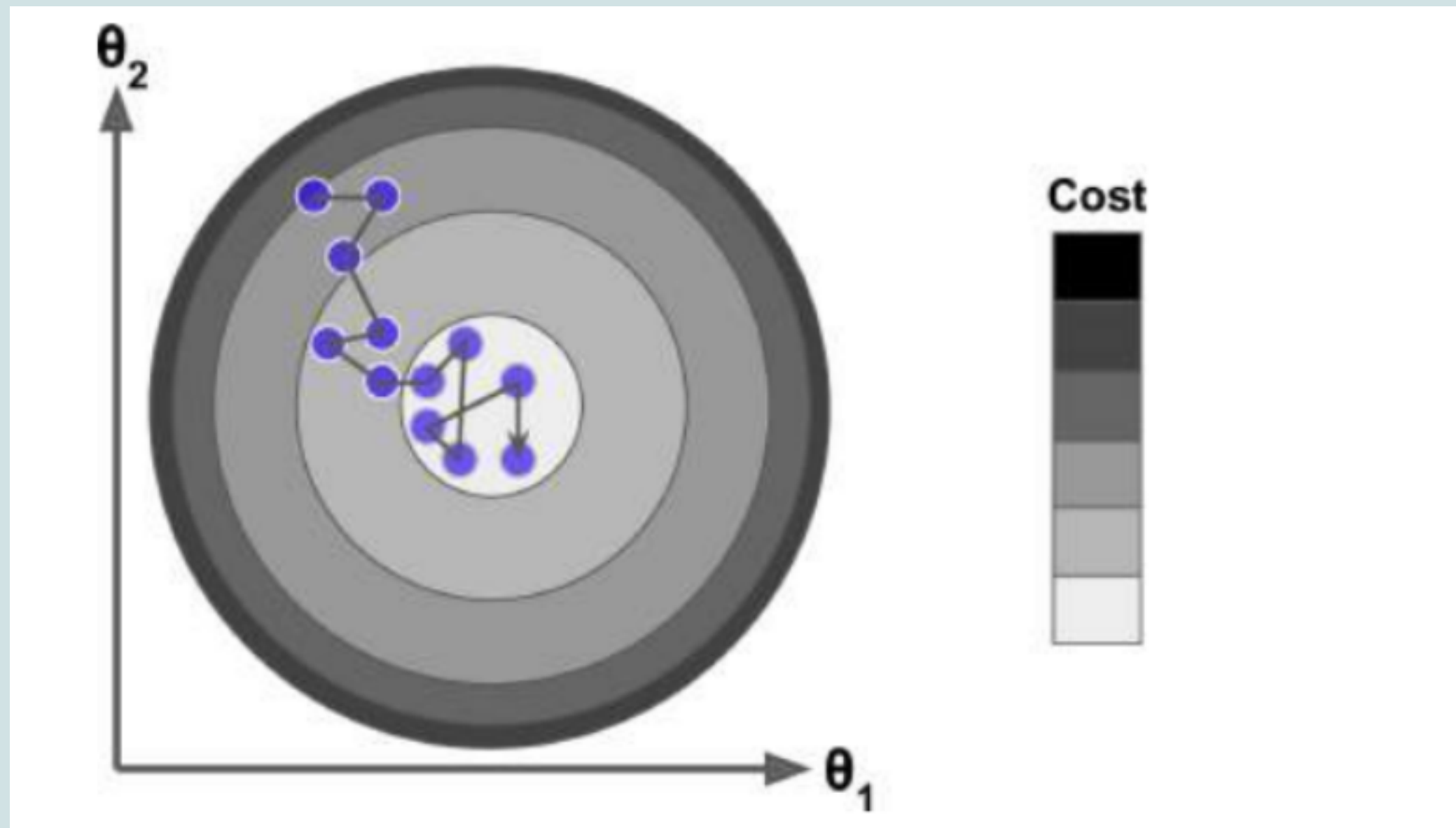
### **Erraticité de la Convergence**

La convergence de la SGD peut être plus erratique en raison de la nature stochastique, présentant des fluctuations dans le progrès vers le minimum.

### **Sensibilité au Learning rate**

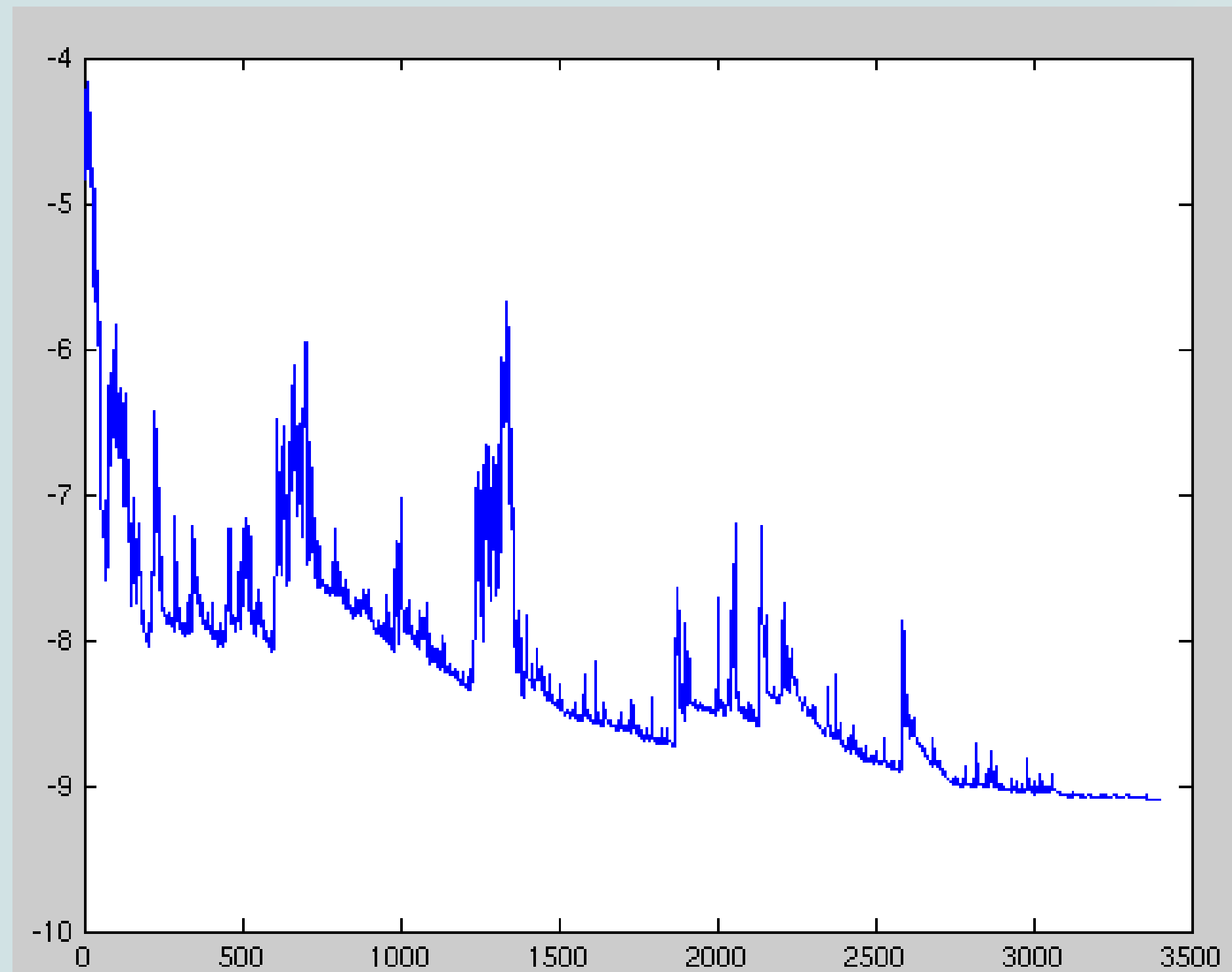
Le choix du learning rate est critique, et une valeur inappropriée peut entraîner des problèmes de convergence.

# Visualisation



La SGD affiche une trajectoire plus irrégulière dans l'espace des paramètres, illustrant sa capacité à converger vers le minimum tout en présentant des fluctuations.

# Visualisation



# Choix du learning rate

La sélection d'un learning rate optimal reste hardie, cela nécessite souvent des techniques d'ajustement dynamique pour garantir la convergence.

En raison de son caractère stochastique, la SGD peut "sauter" hors des minima locaux, offrant un avantage dans la recherche du minimum global.

**V**

**Le mini batch**

# Principe de fonctionnement

La Descente de Gradient par Mini-batch constitue une solution intermédiaire entre la Descente de Gradient par batch et la Descente de Gradient Stochastique, cherchant à allier l'efficacité de la première avec l'adaptabilité de la seconde.

- Elle fonctionne en calculant les gradients sur des ensembles plus petits, appelés mini-lots, qui sont extraits de manière aléatoire de l'ensemble complet des données à chaque itération.
- Ces mini-lots permettent de bénéficier de l'efficacité matricielle tout en atténuant la sensibilité aux anomalies induite par la SGD.



# Avantages et limites

## *Avantages*

### **Efficacité Matricielle**

En traitant des mini-lots, elle profite des optimisations matricielles, notamment lorsque des accélérations matérielles telles que les GPU sont utilisées.

### **Adaptabilité et Réduction de la Variabilité**

Par rapport à la SGD, elle réduit la variabilité du processus tout en conservant une efficacité considérable.

## *Limites*

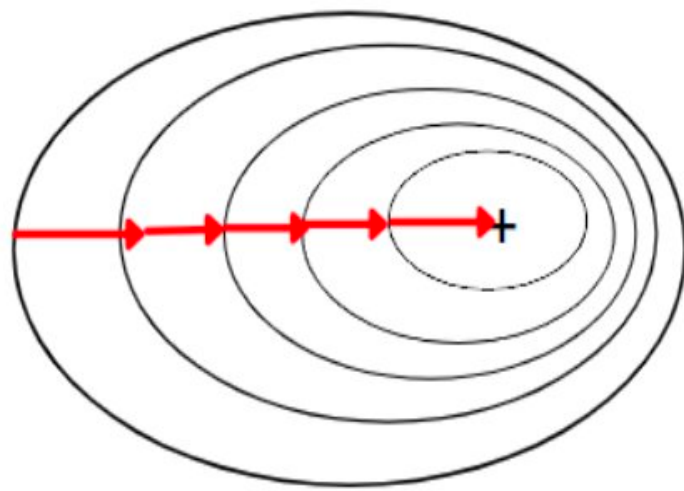
### **Peut Être Sensible au Bruit des Mini-lots**

L'utilisation de mini-lots peut rendre l'algorithme sensible au bruit, surtout avec des mini-lots de petite taille.

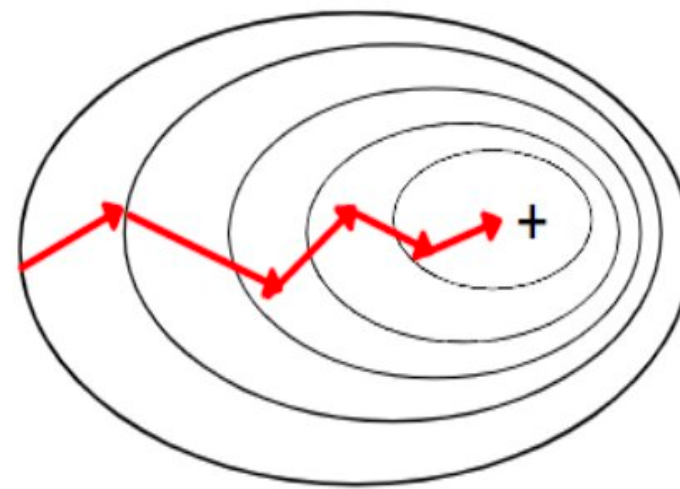


# Visualisation

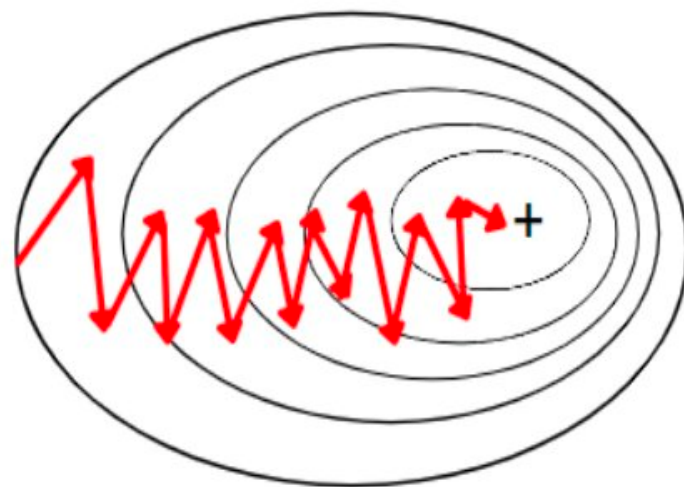
Batch Gradient Descent



Mini-Batch Gradient Descent



Stochastic Gradient Descent



La trajectoire de la Descente de Gradient par Mini-batch dans l'espace des paramètres est plus lisse que celle de la SGD, montrant une convergence plus stable vers le minimum global.

# Choix du learning rate

Comme pour la SGD, le choix d'un taux d'apprentissage approprié demeure essentiel pour garantir une convergence rapide et stable.

La Descente de Gradient par Mini-batch représente un compromis entre l'efficacité de la Descente de Gradient par batch et la stabilité de la Descente de Gradient Stochastique.

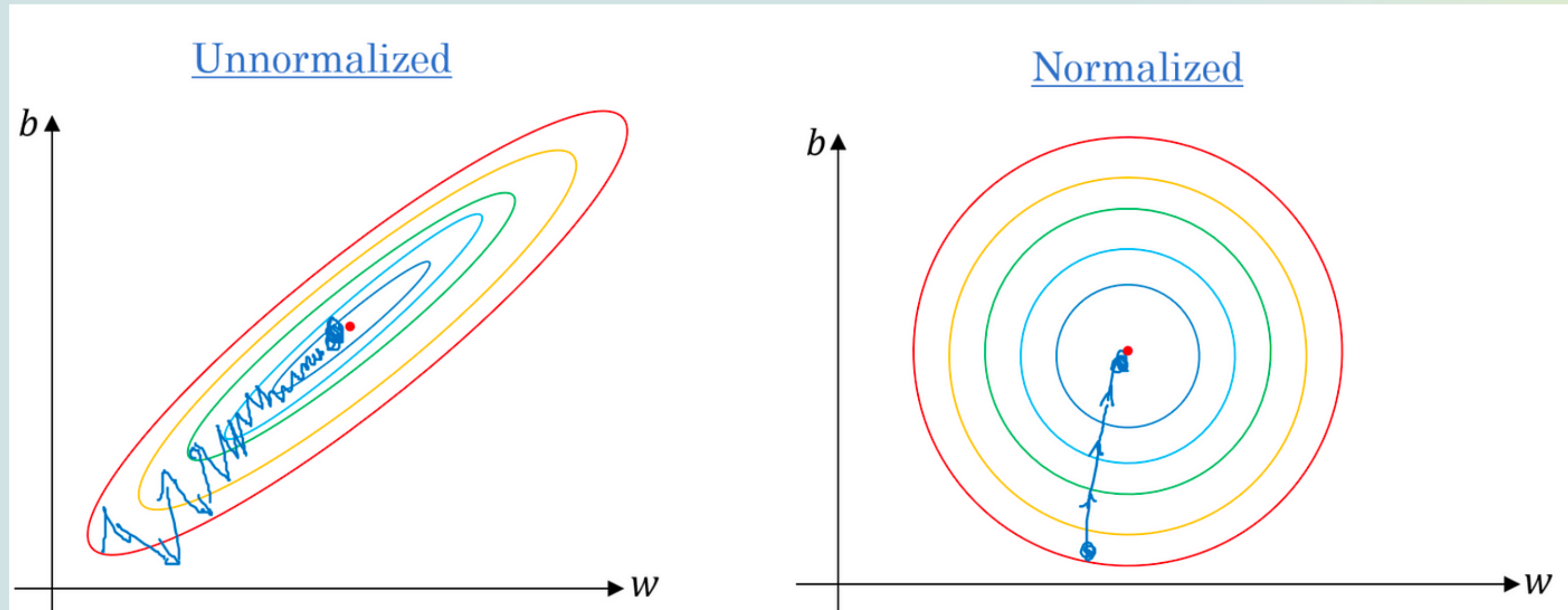
# VI

## Descente du gradient vs Descente du gradient stochastique

# Comparaison des méthodes

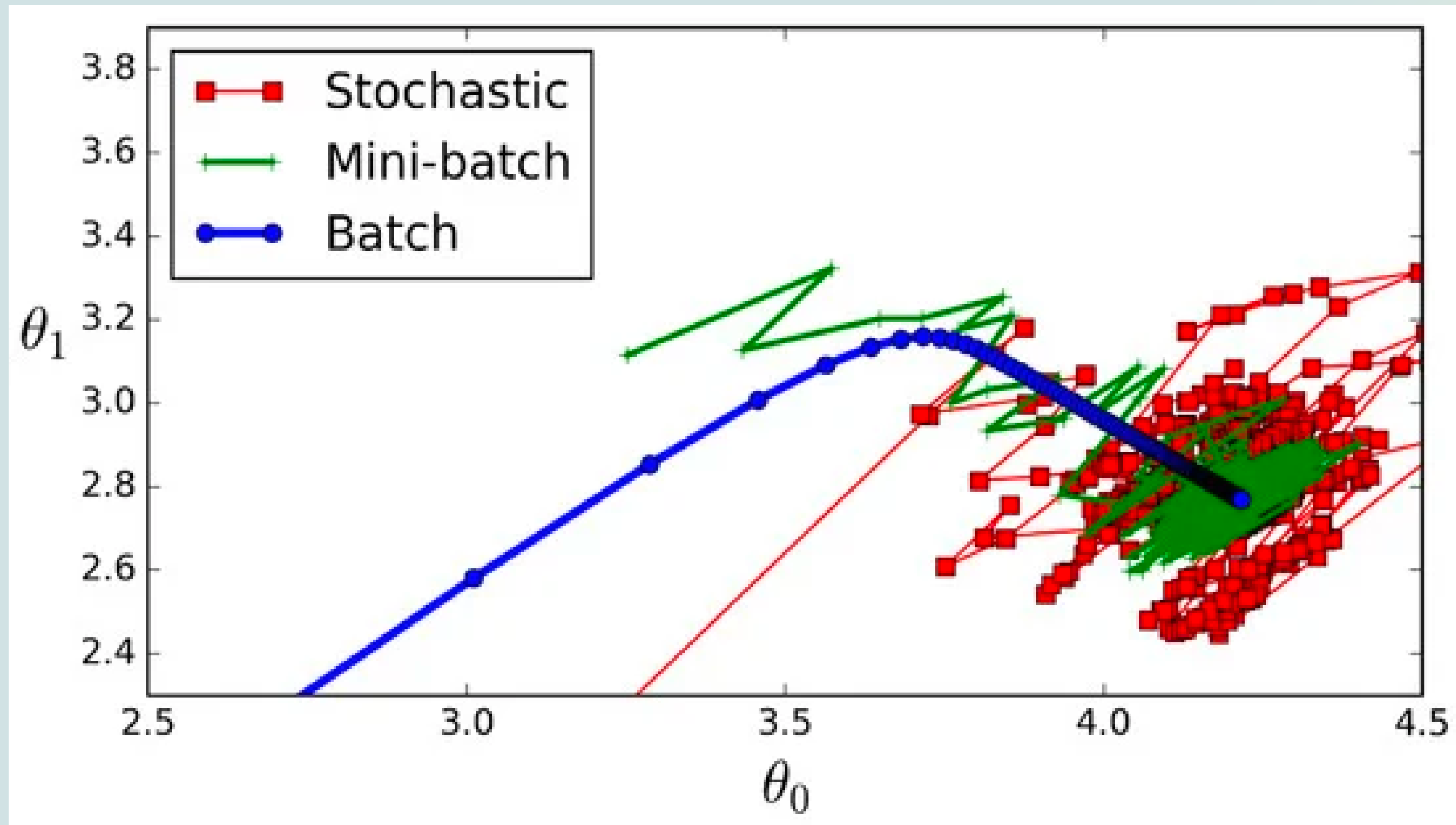
	<b>Descente du gradient stochastique</b>	<b>Batch</b>	<b>Mini batch</b>
Calcul de l'erreur	Pour chaque exemple	Pour tous les exemples	Pour chaque mini batch
Mise à jour du modèle	Pour chaque exemple	Après évaluation de l'ensemble des données	Pour chaque mini batch
Vitesse de convergence	Rapide	Lente	Intermédiaire
Utilisation	Robuste	Efficace	Robuste et efficace

# Visualisation





# Visualisation



# **VII**

## **Conclusions**



La Descente de Gradient par Batch se distingue par sa convergence assurée vers le minimum global, mais elle peut s'avérer coûteuse en calculs, limitant son efficacité sur de vastes ensembles de données.

La Descente de Gradient Stochastique, introduisant l'élément stochastique, excelle sur des ensembles massifs en calculant les gradients sur des instances individuelles. Cependant, sa convergence peut être erratique, nécessitant une gestion minutieuse du taux d'apprentissage.

Enfin, la Descente de Gradient par Mini-batch offre un équilibre judicieux entre l'efficacité matricielle de la Descente de Gradient par batch et la stabilité de la SGD. En travaillant avec des mini-lots, elle optimise le compromis entre l'efficacité calculatoire et la réduction de la variabilité, offrant une alternative équilibrée.