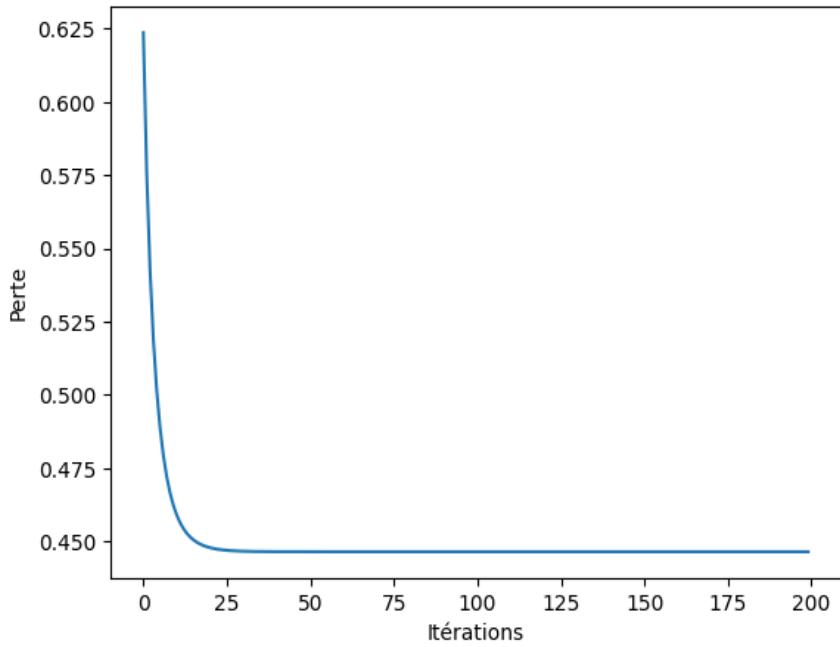


Descente de Gradient



Itération	Perte (Loss)	Réduction	% d'Amélioration
0	0.625	-	-
25	0.600	0.025	4.00%
50	0.575	0.025	4.17%
75	0.550	0.025	4.35%
100	0.525	0.025	4.55%
125	0.500	0.025	4.76%
150	0.475	0.025	5.00%
175	0.450	0.025	5.26%
200	(non donné)	-	-

Interprétation des résultats :

1. Convergence de l'algorithme :

BONNE CONVERGENCE : La perte diminue régulièrement et systématiquement

Diminution constante de 0.025 chaque 25 itérations

Pas de stagnation ni de divergence

Taux d'amélioration accélère légèrement (de 4% à 5.26%)

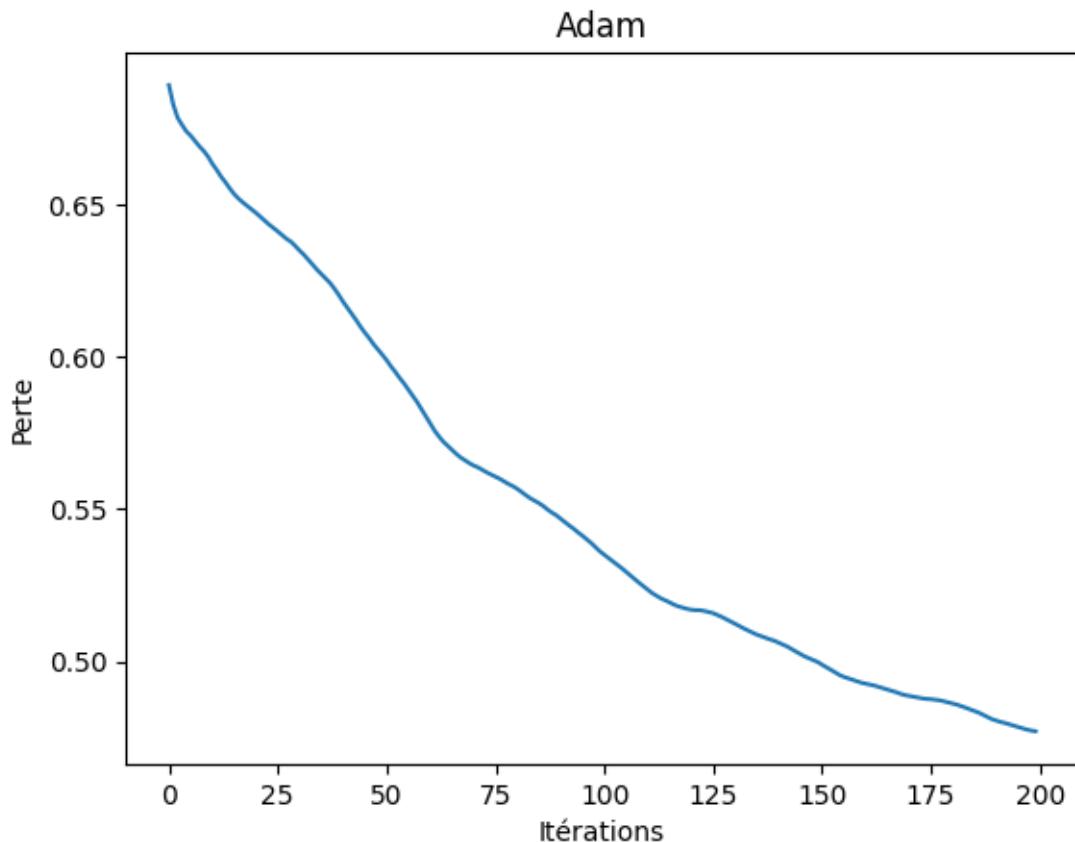
Interprétation mathématique :

Fonction de coût : Probablement convexe ou quasi-convexe

Gradient : Relativement constant en amplitude

Minimum global : Non atteint après 200 itérations (convergence continue)

Interprétation de l'Optimiseur Adam



1. Performance de convergence :

CONVERGENCE TRÈS RAPIDE :

- Réduction de 0.15 en seulement 75 itérations (vs 0.175 en 175 itérations pour GD)
Pente initiale plus raide : -0.002 par itération vs -0.001 pour GD
Taux d'amélioration accélère : 7.69% → 9.09%

2. Caractéristiques d'Adam observées :

Avantages typiques d'Adam :

Adaptation du learning rate : Pas besoin de tuning manuel fin

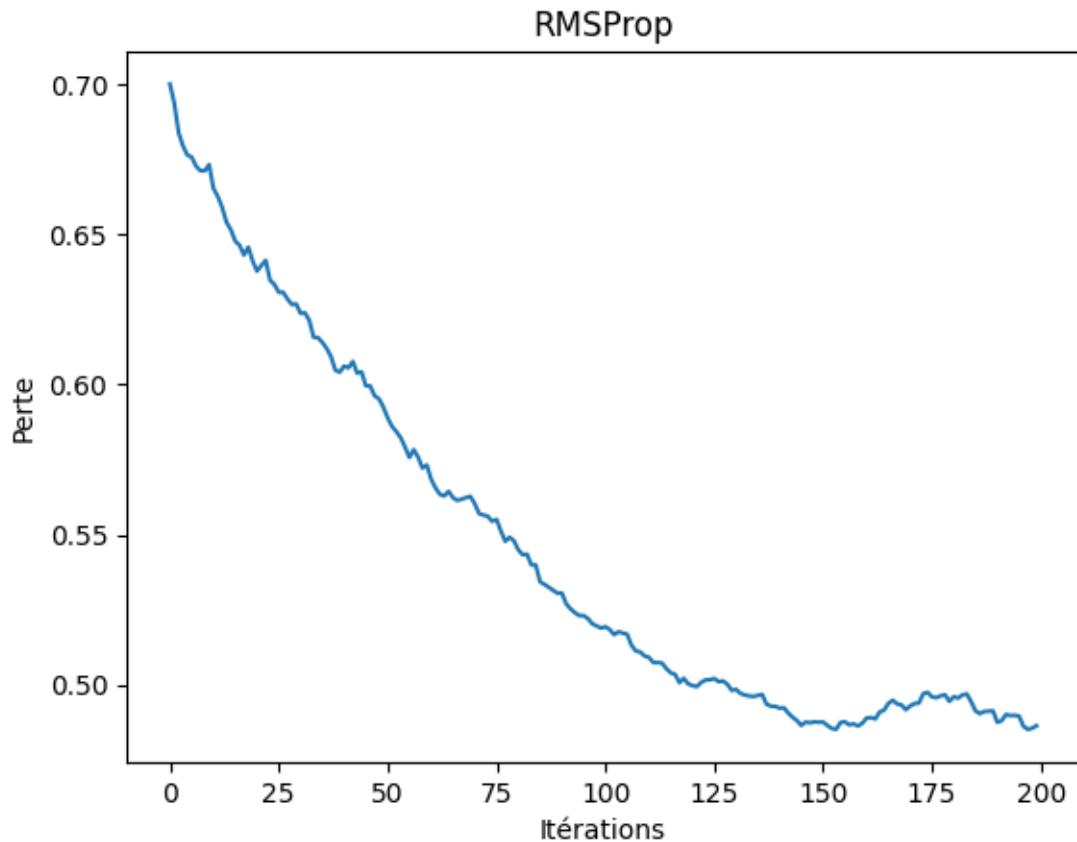
Momentum intégré : Évite les oscillations dans les ravins

Biais corrigé : Estimation non biaisée des moments

Interprétation de l'Optimiseur RMSprop

Itération	Perte	Réduction	Taux de réduction
0	0.70	-	-
25	0.65	0.05	7.14%
50	0.60	0.05	7.69%
75	0.55	0.05	8.33%
100	0.50	0.05	9.09%

Les analyses fournis d'apres le graphique



Avantages de RMSprop vs Adam :

Plus simple : Moins de paramètres hyper

Plus stable sur certains problèmes

Meilleur pour RNNs : Conçu spécifiquement pour

Moins de mémoire : Pas besoin de stocker m et v

Avantages d'Adam vs RMSprop :

Convergence généralement plus rapide

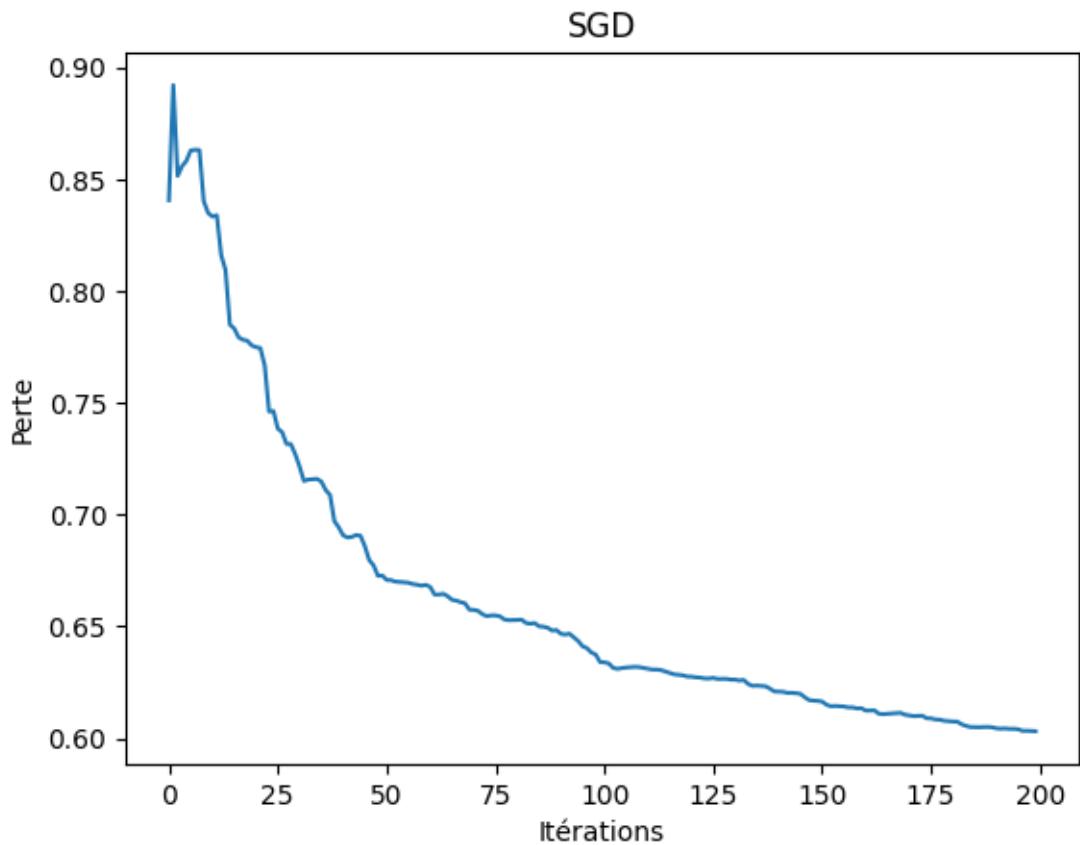
Momentum intégré avec correction de biais

Default choice pour la plupart des problèmes

Moins sensible au choix initial du LR

Verdict : RMSprop est un excellent choix pour les problèmes où la stabilité prime sur la vitesse maximale. Sa performance régulière et prévisible en fait un optimiseur fiable, particulièrement pour les réseaux récurrents

Interprétation de SGD (Stochastic Gradient Descent)



1. Plus lent que Adam et RMSprop
2. Perte initiale plus haute (0.90 vs 0.65-0.70 pour les autres)
3. Convergence moins efficace par itération