

méthode actuelle la plus populaire est appelée Adam, qui est une méthode faisant varier le taux d'apprentissage.

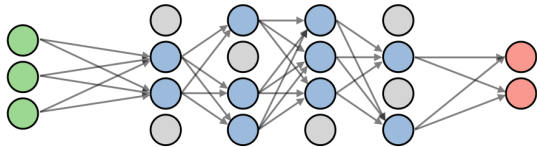
□ **Taux d'apprentissage adaptatifs** – Laisser le taux d'apprentissage varier pendant la phase d'entraînement du modèle peut réduire le temps d'entraînement et améliorer la qualité de la solution numérique optimale. Bien que la méthode d'Adam est la plus utilisée, d'autres peuvent aussi être utiles. Les différentes méthodes sont récapitulées dans le tableau ci-dessous :

Méthode	Explication	Mise à jour de w	Mise à jour de b
Momentum	<ul style="list-style-type: none">- Amortit les oscillations- Amélioration par rapport à la méthode SGD- 2 paramètres à régler	$w - \alpha v_{dw}$	$b - \alpha v_{db}$
RMSprop	<ul style="list-style-type: none">- Root Mean Square propagation- Accélère l'algorithme d'apprentissage en contrôlant les oscillations	$w - \alpha \frac{dw}{\sqrt{s_{dw}}}$	$b \leftarrow b - \alpha \frac{db}{\sqrt{s_{db}}}$
Adam	<ul style="list-style-type: none">- Adaptive Moment estimation- Méthode la plus populaire- 4 paramètres à régler	$w - \alpha \frac{v_{dw}}{\sqrt{s_{dw}} + \epsilon}$	$b \leftarrow b - \alpha \frac{v_{db}}{\sqrt{s_{db}} + \epsilon}$

Remarque : parmi les autres méthodes existantes, on trouve Adadelata, Adagrad et SGD.

3.4 Régularisation

□ **Dropout** – Le dropout est une technique qui est destinée à empêcher le sur-ajustement sur les données de training en abandonnant des unités dans un réseau de neurones avec une probabilité $p > 0$. Cela force le modèle à éviter de trop s'appuyer sur un ensemble particulier de features.



Remarque : la plupart des frameworks d'apprentissage profond paramètrent le dropout à travers le paramètre 'garder' $1 - p$.

□ **Régularisation de coefficient** – Pour s'assurer que les coefficients ne sont pas trop grands et que le modèle ne sur-ajuste pas sur le training set, on utilise des techniques de régularisation sur les coefficients du modèle. Les techniques principales sont résumées dans le tableau suivant :