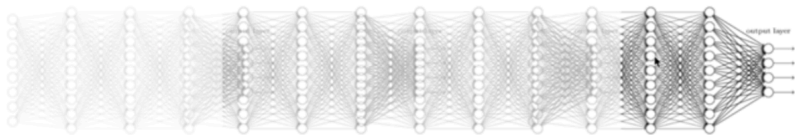


Deep Learning par la Pratique

Optimisation/Régularisation de réseaux profonds

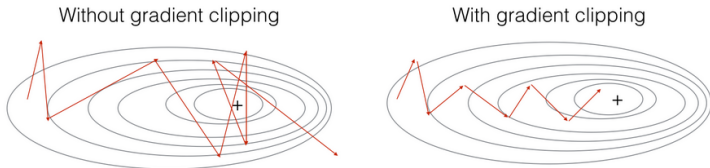
Optimisation de réseaux profonds

Disparition du gradient (ou son explosion)



Optimisation de réseaux profonds

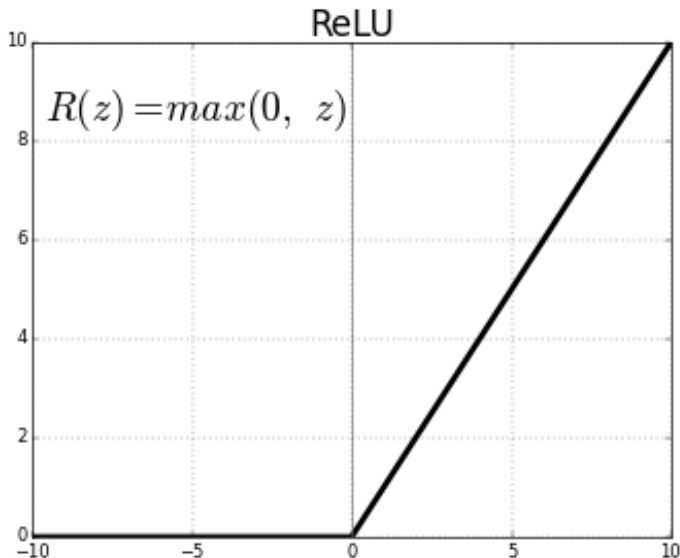
Exploding Gradient \Rightarrow Gradient clipping



Dans le cadre des réseaux récurrents \Rightarrow initialisation avec des matrices orthogonales

Optimisation de réseaux profonds

Vanishing Gradient \Rightarrow utilisation de ReLu plutôt que les sigmoïdes



Optimisateur plus rapide que SGD

- Root Mean Square Propagation (RMSProp)
- Adaptative Gradient Algorithm (AdaGrad)
- Adaptative Moment Estimation (Adam) \Leftarrow
- ... AdaBound (2019) ?