

# Adaptive learning rates

## Ada grad

$$\omega^n = \omega^{n-1} - \frac{2}{\sqrt{R} + \epsilon} g^n$$

$$R = \sum_i (g^n)^2$$

## RMS prop

$$\omega^n = \omega^{n-1} - \frac{2}{\sqrt{R^n + \epsilon}} g^n$$

$$R^n = \rho R^{n-1} + (1-\rho)(g^n)^2$$

2014

Adam

→ (RMSprop) + (SGD with momentum)

SGD with momentum