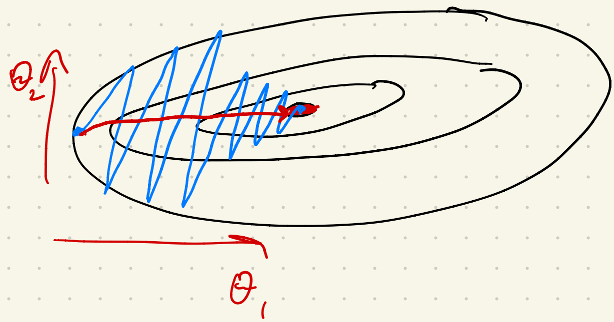
momentum $\quad 2n + k$

$$V_i = \beta V_{i-1} + \nabla_\theta L(x)$$

$$\theta_i = \theta_{i-1} - \gamma V_i$$



---

rms prop $\quad 2n + k$

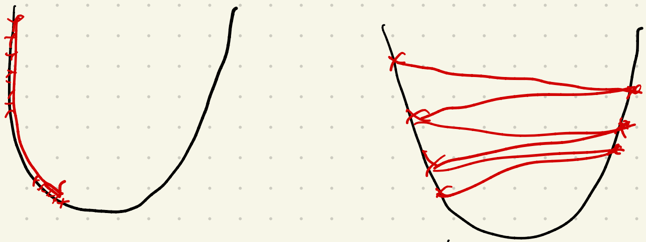
$$V_i = \beta V_{i-1} + \nabla_\theta L(x)^2 (1-\beta)$$

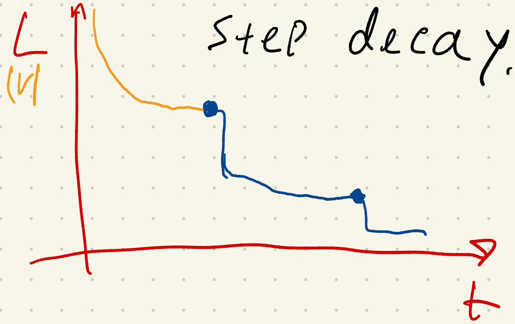
$$\theta_i = \theta_{i-1} - \gamma \frac{\nabla_\theta L(x)}{\sqrt{V_i} + \varepsilon}$$

---

adam $\quad 2n + 2k$

γ - learning rate. γ - ?





AmsGrad.

$$\left[\hat{V}_i = \max(V_i, V_{i-1})\right]$$

torch. clamp(X, ~~lower~~, upper).

$$\boxed{\theta_i \sim N(1, 2)} \quad \forall i \qquad \theta_i \sim N(1, 0.6)$$

$$\underbrace{((x^T \tilde{\theta}_1) \times \tilde{\theta}_2) \times \ldots \times \tilde{\theta}_n}_{} = t \qquad \begin{pmatrix} 1.5 & 0 \\ 0 & 1.5 \end{pmatrix}$$

$$t = \begin{pmatrix} 1.5 & 0 \\ 0 & 1.5 \end{pmatrix}^n$$

$$\begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix}^n$$

$$\theta_i \sim N(1, 1)$$

$$t = \theta^T x + b = \left[\sum_{i=1}^{n} \theta_i x_i\right]$$

1) θ_i iid

2) x_i iid

3) $\theta_i \perp x_i$

$$V(t_i) = \mathbb{E}(\theta_i x_i)^2 - (\mathbb{E}\theta_i x_i)^2 =$$

$$= \mathbb{E}\theta_i^2 \mathbb{E}x_i^2 - (\mathbb{E}\theta_i x_i)^2 =$$

$$= \left(\mathbb{E}\theta_i^2 - (\mathbb{E}\theta_i)^2 + (\mathbb{E}\theta_i)^2\right)\left(\mathbb{E}x_i^2 - (\mathbb{E}x_i)^2 + (\mathbb{E}x_i)^2\right)$$

$$- (\mathbb{E}\theta_i x_i)^2 = \left(Var(\theta_i) + (\mathbb{E}\theta_i)^2\right) \times$$

$$\times \left(Var(x_i) + (\mathbb{E}x_i)^2\right) - (\mathbb{E}\theta_i x_i)^2 =$$

$$\mathbb{E}\theta_i = 0$$

$$\mathbb{E}x_i = 0$$

$$= Var(\theta_i) \cdot Var(x_i) = Var(t_i).$$

$$Var(t) = \sum_{i=1}^{k} Var(t_i) = \underline{k}\, Var(\theta_i)\, Var(x_i).$$

$\theta_i \sim N(0, 1).$

$k = 2.$ $n = 32.$

$Var(\theta_i) = 1$ $Var(x_i) = 1.$

$Var(t_{out}) = 2^{32}$

$$\theta_i \sim U\left[\frac{-1}{\sqrt{k}}, \frac{1}{\sqrt{k}}\right]$$

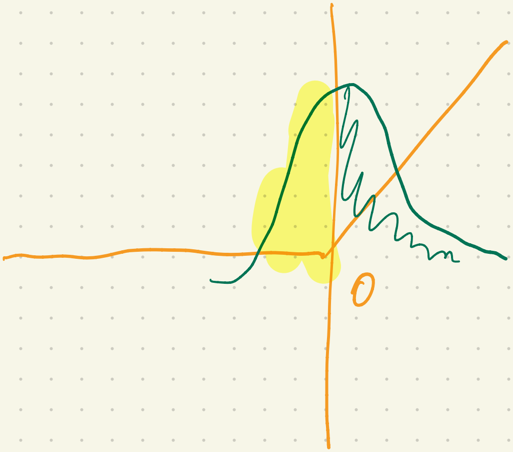
$$Var(\theta_i) = \boxed{\frac{1}{3k}}$$

$Var(t) =$
$= k \cdot \frac{1}{3 \cdot k} \cdot 1. =$
$= \frac{1}{3}.$

$$Var(\theta_i) = \frac{2}{k+m}.$$

Glorot init
Xavier init.

$$\theta_i \sim U\left[\frac{-\sqrt{6}}{\sqrt{k+m}}, \frac{\sqrt{6}}{\sqrt{k+m}}\right].$$



He. Kaiming.

$$\theta_i \sim N\left(0, \sqrt{\frac{2}{m}}\right).$$