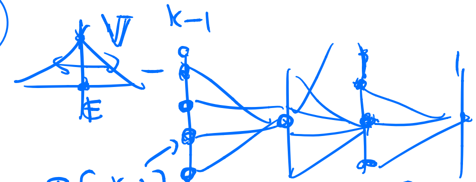
1.

$$a_j = \sum_i W_{ij}^k x_i^{k-1}$$



$$\mathbb{E}[a_j] = \sum_i \mathbb{E}(W_{ij}^k x_i^{k-1})$$

$$= \sum_i \mathbb{E}[W_{ij}^k] \cdot \mathbb{E}(x_i^{k-1})$$

$$\overset{\shortparallel}{0}.$$

$$\mathbb{E}[x_i^{k-1}] = \mathbb{E}[x^{k-1}]$$

$$\mathbb{V}[x_i^{k-1}] = \mathbb{V}[x^{k-1}]$$

$$\mathbb{E}[a] = 0$$

$$\mathbb{E}[W_{ij}^k] = \mathbb{E}[W^k] = 0$$

$$\mathbb{V}[W^k] \in \mathbb{R}.$$

$$\mathbb{V}[a] = \sum_i \mathbb{V}[W_{ij}^k x_i^{k-1}]$$

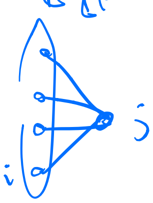
$$\mathbb{V}[XY] = \mathbb{V}[X] \cdot \mathbb{E}[Y^2] + \mathbb{V}[Y] \cdot \mathbb{E}[X^2] + \mathbb{V}[X] \cdot \mathbb{V}[Y]$$

$$\mathbb{V}[W_{ij}^k x_i^{k-1}] = \mathbb{E}[(W_{ij}^k)^2 (x_i^{k-1})^2]$$

$$= \underbrace{\mathbb{E}[(W_{ij}^k)^2]}_{\mathbb{V}[W^k]} \cdot \underbrace{\mathbb{E}[(x_i^{k-1})^2]}_{\mathbb{E}[(x^{k-1})^2]}$$

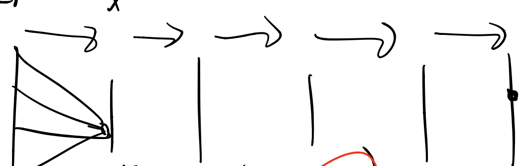
$$\mathbb{V}[a_j] = \sum_i \mathbb{V}[W^k]\, \mathbb{E}[(x^{k-1})^2]$$

$$n^{k-1} \cdot \mathbb{V}[W^k] \cdot \mathbb{E}[(x^{k-1})^2] .$$

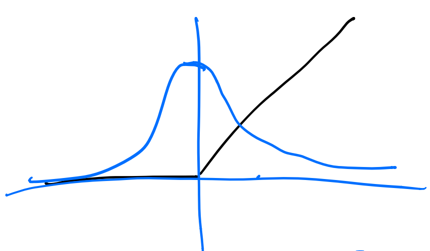


$$\mathbb{E}[(x^k)^2]$$

x^{k-1} x^k



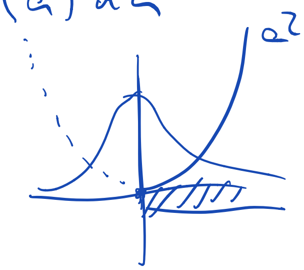
$$W^k \sim N(0, \sigma^2)$$



$$\mathbb{E}[\max(a, 0)]$$

$$\mathbb{V}[\max(a, 0)]$$

$$\mathbb{E}[\max(e,0)^2] = \int_{-\infty}^{\infty} \underline{\max(e^2,0)} \cdot p(e)\, de$$



$$= \int_{0}^{\infty} e^2 p(e)\, de$$

$$= \frac{1}{2}\int_{-\infty}^{\infty} e^2 p(e)\, de = \frac{1}{2}\mathbb{E}[e^2]$$

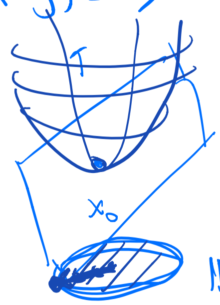
$$= \frac{1}{2}\mathbb{V}[e]$$

$$\mathbb{E}[(x^k)^2] = \underbrace{\frac{1}{2} n_{k-1} \mathbb{V}[W]}_{= 1} \mathbb{E}[(x^{k-1})^2]$$

$$\to \sigma^2 = \frac{2}{n_{k-1}}$$

③. $\min \langle g, \Delta x\rangle + \underline{\langle \Delta x, M\Delta x\rangle}$ $\quad \nabla f(x_0) = g$

$M \succ 0$



$\|\Delta x\|_M \leq \varepsilon$.

- $\min_{\Delta x} \langle g, \Delta x\rangle$ $\quad ⊜ \quad \min_{\Delta x} \max_{\lambda \geq 0} \langle g, \Delta x\rangle + \lambda \cdot (\|\Delta x\|_M^2 - \varepsilon)$ $\quad > 0$.

$(\underline{\|\Delta x\|_M^2 \leq \varepsilon^2})$ $\quad \cdot \quad \lambda$

- $L(\Delta x, \lambda) = \langle g, \Delta x\rangle + \lambda(\|\Delta x\|_M^2 - \varepsilon^2)$

$$0 = \frac{\partial}{\partial \Delta x} = g + 2\lambda M \Delta x \qquad \langle \Delta x, M \Delta x \rangle$$

$$\Delta x = \underline{-\frac{1}{2\lambda} M^{-1} g}$$

$$\|\Delta x\|_M^2 = \varepsilon^2$$

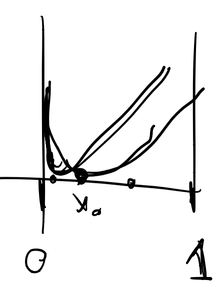
$$\left\langle -\frac{1}{2\lambda} M^{-1} g, M \frac{-1}{2\lambda} M^{-1} g \right\rangle = \varepsilon^2$$

$$\frac{1}{4\lambda^2} \langle M^{-1} g, g \rangle = \varepsilon^2$$

$$\lambda = \frac{1}{2\varepsilon} \sqrt{\langle g, M^{-1} g \rangle}$$

$$\Delta x = -\varepsilon \frac{1}{\sqrt{\langle g, M^{-1} g \rangle}} \cdot M^{-1} g$$

(4). $\min_{x} \langle g, x - x_0 \rangle + \lambda D(x, x_0)$



- $D(x, x_0) = x \log \frac{x}{x_0} + (1-x) \log \frac{1-x}{1-x_0}$

$$0 = \frac{\partial}{\partial x} = g + \lambda \left(\log \frac{x}{x_0} + \underbrace{x \cdot \frac{1}{x}}_{1} \right.$$

$$\left. - \log \frac{1-x}{1-x_0} - \underbrace{(1-x) \cdot \frac{1}{1-x}}_{1} \right)$$

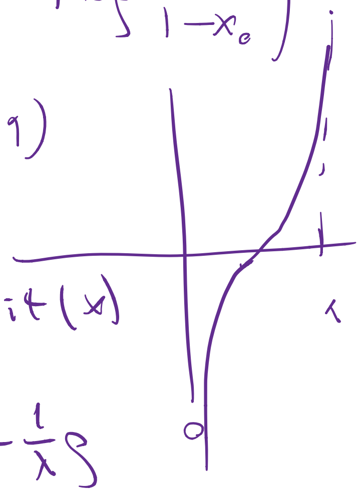
$$0 = g + \lambda \left(\log \frac{x}{x_0} - \log \frac{1-x}{1-x_0} \right)$$

$$0 = g + \lambda\left(\log\frac{x}{1-x} - \log\frac{x_0}{1-x_0}\right)$$

$$x \in (0, 1)$$

$$x = \sigma(y)$$

$$y = \log\frac{x}{1-x} = \operatorname{logit}(x)$$



$$\operatorname{logit}(x) = \operatorname{logit}(x_0) - \frac{1}{\lambda} g$$

$$y_{t+1} = y_t - \lambda g$$

$$x_{t+1} = \frac{1}{1+e^{-y_{t+1}}} = S(y_{t+1})$$

$$x = S(\operatorname{logit}(x_0) - \lambda g).$$