$$f_\theta(x) = \sum_{i=1}^{m} \alpha_1 a_i \,\sigma(\alpha_2 w_i \cdot x)$$

$$a_i(0) \sim N(0, \beta_1^2)$$

$$w_{ij}(0) \sim N(0, \beta_2^2)$$

$$\tilde{\theta} = \begin{bmatrix} [a_i/\beta_1]_i \\ [w_{ij}/\beta_2]_{ij} \end{bmatrix}$$

$g_\theta(x)$

$$f_\theta(x) = \sum_{i=1}^{m} \alpha_1 \beta_1 \tilde{a}_i \,\sigma(\alpha_2 \beta_2 \tilde{w}_i \cdot x) \quad \text{ReLU}$$

$$= \alpha_1 \alpha_2 \beta_1 \beta_2 \Big(\underbrace{\sum_{i=1}^{m} \tilde{a}_i \,\sigma(\tilde{w}_i \cdot x)}_{\tilde{f}_{\tilde{\theta}}(x)} \Big)$$

$$\dot{\theta} = -\sum_j \nabla_\theta f_\theta(x_j) \left[f_\theta(x_j) - f(x_j) \right]$$

$$\left[f_\theta(x_j) - f(x_j) \right] = \alpha_1\alpha_2\beta_1\beta_2 \left[\tilde{f}_{\tilde{\theta}}(x_j) - \frac{1}{\alpha_1\alpha_2\beta_1\beta_2} f(x_j) \right]$$

$$\nabla_\theta f_\theta(x_j) = \begin{bmatrix} [\alpha_1 \sigma(\alpha_2 w_i \cdot x_j)]_i \\ [\alpha_1 a_i \sigma'(\alpha_2 w_i \cdot x_j)\, \alpha_2 x_j]_i \end{bmatrix}$$

$$= \alpha_1 \alpha_2 \begin{bmatrix} \beta_2 [\sigma(\tilde{w}_i \cdot x_j)]_i \\ \beta_1 [\tilde{a}_i \sigma'(\tilde{w}_i \cdot x_j) x_j]_i \end{bmatrix}$$

$$\frac{d\tilde\theta}{dt} = -\sum_j \begin{bmatrix} \beta_2/\beta_1 I_m & \\ & \beta_1/\beta_2 I_{m\cdot d} \end{bmatrix} \alpha_1^2\alpha_2^2\beta_1\beta_2 \nabla_{\tilde\theta} \hat{f}_{\tilde\theta}(x_j) \left[\hat{f}_{\tilde\theta}(x_j) - \frac{1}{\alpha_1\alpha_2\beta_1\beta_2} f(x_j)\right]$$

$\boxed{\tilde t = \alpha_1^2\alpha_2^2\beta_1\beta_2 t}$

$\tilde f$

$$\frac{d\tilde\theta}{d\tilde t} = -\sum_j \begin{bmatrix} \beta_2/\beta_1 I_m & \\ & \beta_1/\beta_2 I_{md} \end{bmatrix} \nabla_{\tilde\theta} \tilde{f}_{\tilde\theta}(x_j) \left[\tilde{f}_{\tilde\theta}(x_j) - \frac{1}{\alpha_1\alpha_2\beta_1\beta_2} f(x_j)\right]$$

$\boxed{y = \sqrt{m^2}\, y_0}$

$$\frac{d\tilde\theta}{dt} = -\sum_j \nabla_{\tilde\theta} \hat{f}_{\tilde\theta}(x_j) \left[\hat{f}_{\tilde\theta}(x_j) - \hat{f}(x_j)\right]$$

$K(x, x')$

$\mathbb{E}_{\rho_0}\left[A(x,\theta) A(x',\theta)\right]$

$\downarrow$

0

$$K = \frac{1}{\alpha\beta_1\beta_2}$$

$\alpha_1\alpha_2$

$$\left[\beta_2/\beta_1,\ \alpha_1^2\alpha_2^2\beta_1\beta_2,\ \frac{1}{\alpha_1\alpha_2\beta_1\beta_2}\right]$$

$\uparrow$ t $\qquad \alpha \qquad \uparrow$

$\delta\rho_t(\theta) = \rho_t(\theta) - \rho_0(\theta)$

continuum limit

m finite

$\mathbb{R}^{d+1}$

$K \to 0$

$K \to \infty$

$m \to \infty$

$K \sim m$

$K \sim \sqrt{m}$

$K \sim m^{\gamma}$

(reduced $\theta \in \mathbb{R}^{d+1}$)

$\rho_0 \qquad \rho_{\tilde t}$

∂_t

$$\lim_{m\to\infty} \mu_T^{(m)} - \mu_0 \xrightarrow{?} 0$$

$$\forall f \in C_c \quad \int f\left[d\mu_T^m - d\mu_0\right] \to 0$$

$$\Big[\overbrace{\underbrace{a_i|w_i|}_{\text{amplitude}}}^{r_i},\ \underbrace{\hat{w}_i}_{\text{feature}}\Big]$$

$$f_\theta(x) = \sum_{i=1}^{m} a_i|w_i|\sigma(\hat{w}_i\cdot x)$$

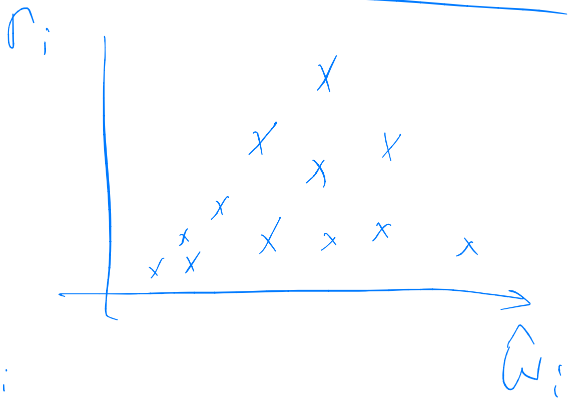
$$\pi(r_i, \hat{w}_i)$$

$$f_\theta(x) \overset{?}{=} m\int a(\hat{w})\,\sigma(\hat{w}\cdot x)\, d\pi(\hat{w})$$

$$a(\hat{w}) = E_i\big(a_i|w_i| \;\big|\; \hat{w}_i = \hat{w}\big)$$

$$f(x) = \sum_{i=1}^{m} r_i\,\sigma(\hat{w}_i\cdot x)$$

$$\pi(r_i, \hat{w}_i)$$



$$\int r_i\,\sigma(\hat{w}_i\cdot x)\boxed{\rho(r_i, \hat{w}_i)}\,dr_i\, d\hat{w}_i$$

$$r_i\,\rho_{\hat{w}_i}(r_i)\,\rho(\hat{w}_i)\,\sigma(\hat{w}_i\cdot x)$$

$$\int r_i\,\rho_{\hat{w}_i}(r_i)\,dr_i$$

$$\frac{d\tilde{\theta}}{dt} = -\sum_j \nabla_{\tilde{\theta}} f_{\tilde{\theta}}(x_j)\big[f_{\tilde{\theta}}(x_j) - k f(x_j)\big]$$

$$f_\theta(x) = m\boxed{\int a\cdot\sigma(w\cdot x)\,\rho(a, |w|, \hat{w})\, da\, d|w|\, d\hat{w}} \sim \infty$$

$$k \sim m : \lim_{m\to\infty} \lim_{\delta t\to 0} \sum_{t=0}^{T} W\big(\rho^{(m)}_{t+\delta t}, \rho^{(m)}_t\big) \sim O(1)$$

$$0$$

$$k \sim \sqrt{m}$$

$$\lim_{m\to\infty}\left[\int \|\dot{\theta}_i(t)\|_2\,dt\right] \sim O(m^\beta)$$

$$\rho(0) \sim N(0,1)$$

$$\rho : \frac{k}{m} f$$

<1> $k = m^\gamma$

$$f_\theta(x) = \frac{1}{k}\sum_{i=1}^{m} a_i\,\sigma(w_i x + b_i)$$

$$a_i, w_i, b_i \sim N(0,1)$$

<2> $\theta = (a_i, w_i, b_i)$ GD

(3) Monitor.

(1) $$\alpha_\gamma^{(m)} = \sum_{t=0}^{T} \frac{1}{m}\sum_{i=1}^{m} \left\| \frac{\theta_{t+1}^i - \theta_t^i}{\Delta t} \right\|_2 \Delta t$$

$$\|V\|_2 = \sqrt{\sum V_i^2}$$

Given tolerance

Give Max epochs T_0

the leaing rate $\eta = \frac{\eta_0}{m}$

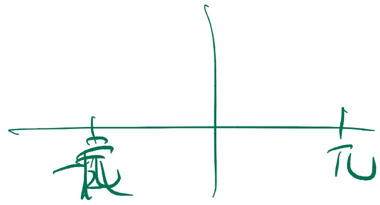
precision should be approx same when m is large.

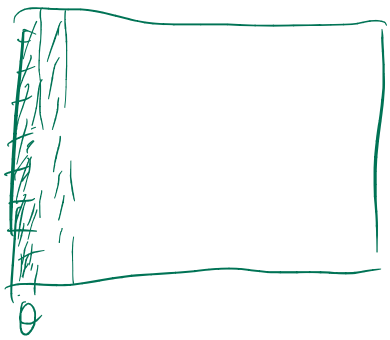
② $\frac{1}{m}\sum_{i=1}^{m} \|\theta^i_{T_0} - \theta^i_0\|_2$

③ Amplitude vs orientation

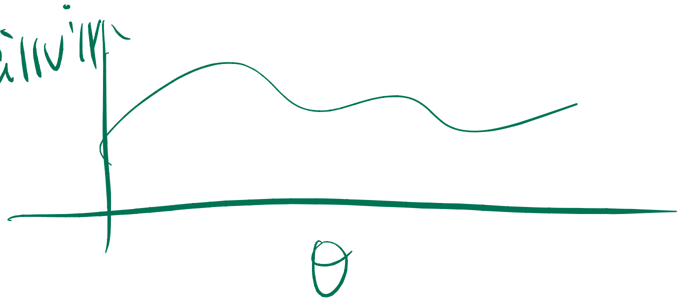
$$v^i \doteq (w^i, b^i)$$

$a^i \|v^i\|$ vs. $\frac{v^i}{\|v^i\|}$ $\quad \theta$





Heatmap of $\|v^i\|$



Study $\gamma = 1$, $\gamma < 1$, $\gamma > 1$

Guess: $\lim_{m\to\infty} Q_\gamma^{(m)} = \begin{cases} 0 & \gamma < 1 \\ O(1) & \gamma = 1 \\ \infty & \gamma > 1 \end{cases}$