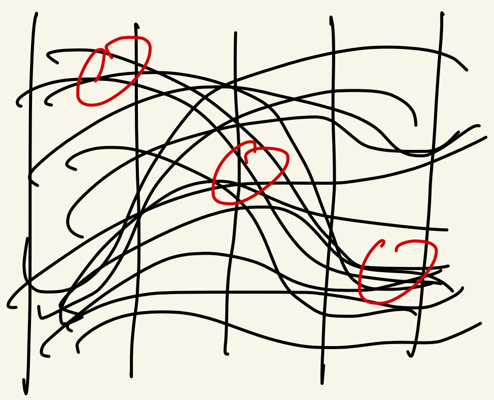
Gaussian Process



$$\bar{x} \sim \mathcal{N}^{D=5}(\bar{\mu} = \bar{0}, K)$$

$$\lim_{D \to \infty} \mathcal{N}^{D}(\bar{\mu}, K) \longrightarrow GP(\mu(x), K(x_i, x_j))$$

→ Distribution over functions

→ Allowed by Mercer theorem

Problem statement

select $f_i \in GP$ s.t. they pass thru training data

$$T = \overbrace{\{ f_i \in GP : f_i(x) = T(x) \}}^{\mu_i, \Sigma \sim GP(\mu_i', K)}$$

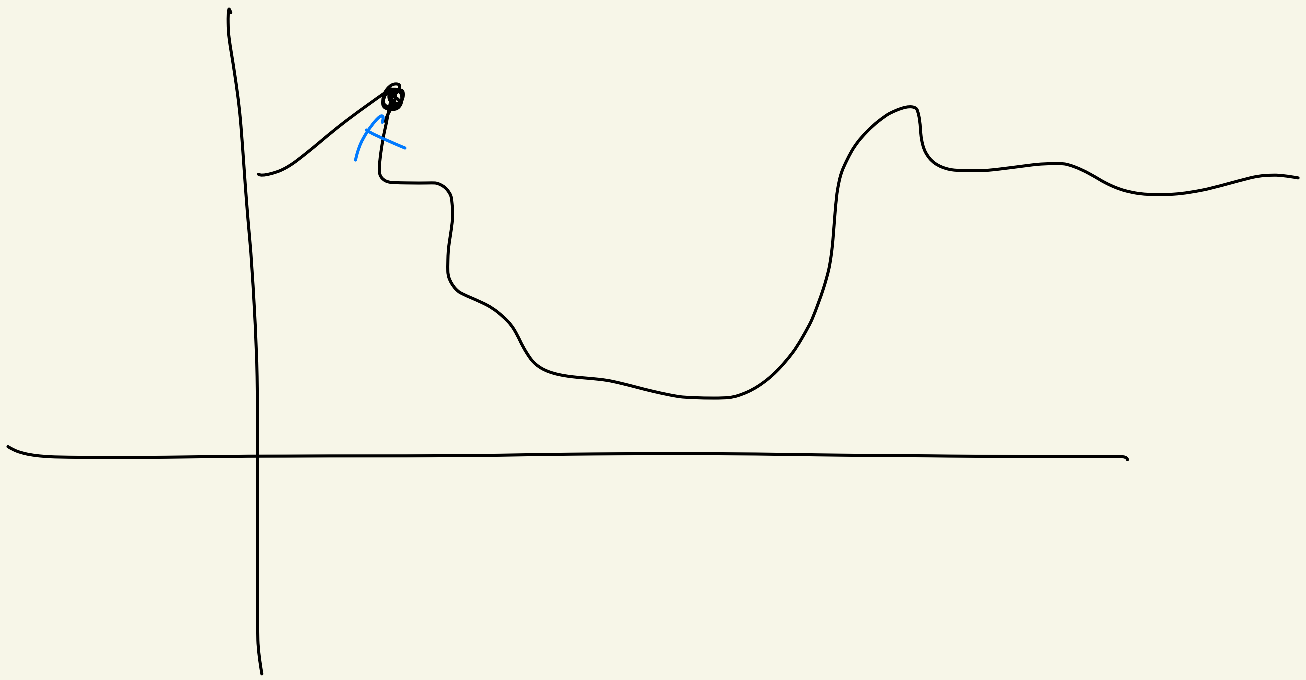
$$\left.\begin{aligned} K(x, X) &= K \\ K(x^*, x^*) &= K_{**} \\ K(x, x^*) &= K_* \end{aligned}\right\}$$

$$y^* = K_*^T K^{-1} y$$

$$\Sigma^* = K_{**} - K_*^T K^{-1} K_*$$

$\rightarrow$ Adding Noise

$$K \rightarrow K + \sigma^2 I$$



$$y^* = K_*^T (K + \sigma^2 I)^{-1} y$$

$$\Sigma^* = K_{**} - K_*^T (K + \sigma^2 I)^{-1} K_*$$

→ Performance

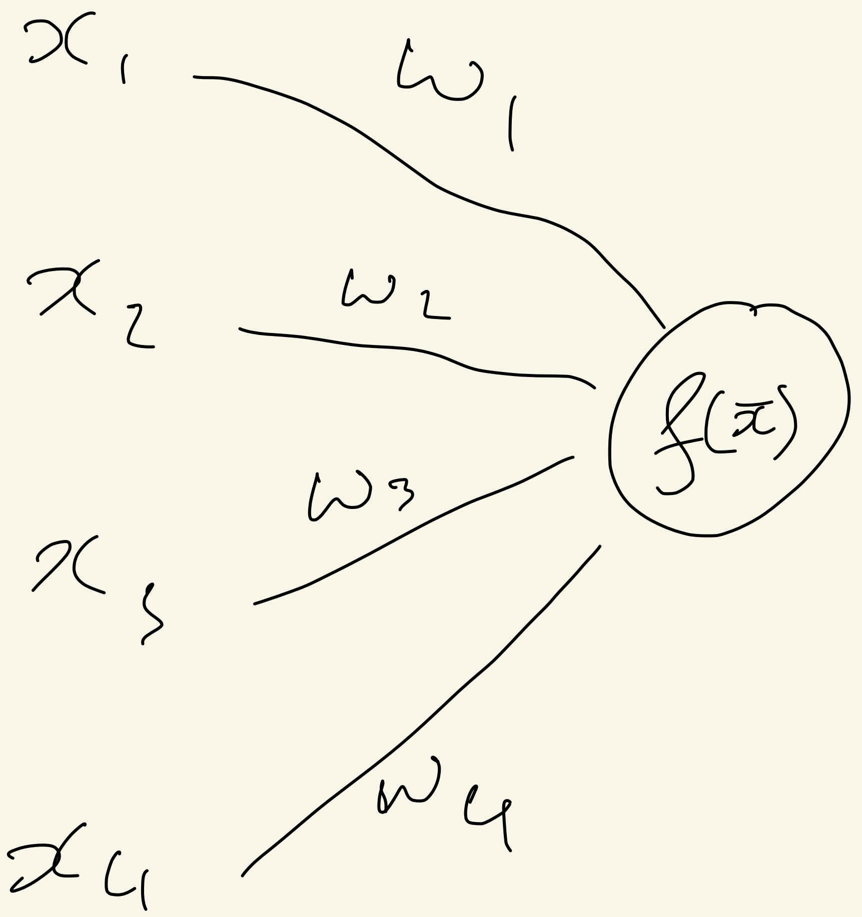
$$y_* = \overbrace{k_*^T}^{\text{only new thing}} \underbrace{(K + \sigma^2 I)^{-1}}_{O(n^3) \to O(n^{2.373})} y$$

$$\therefore y_* = \underbrace{k_*^T \underline{\alpha}}_{O(n)} \quad : \quad \underline{\alpha} = (K + \sigma_n^2 I)^{-1} y$$

→ Sparse - GPR

↳ If time

Perceptron Theory



$$= \begin{cases} 1, & w_i x_i > S \\ 0, & ow \end{cases}$$

Activation function.

1. $\bar{x} \to f \to y^{Pred}$

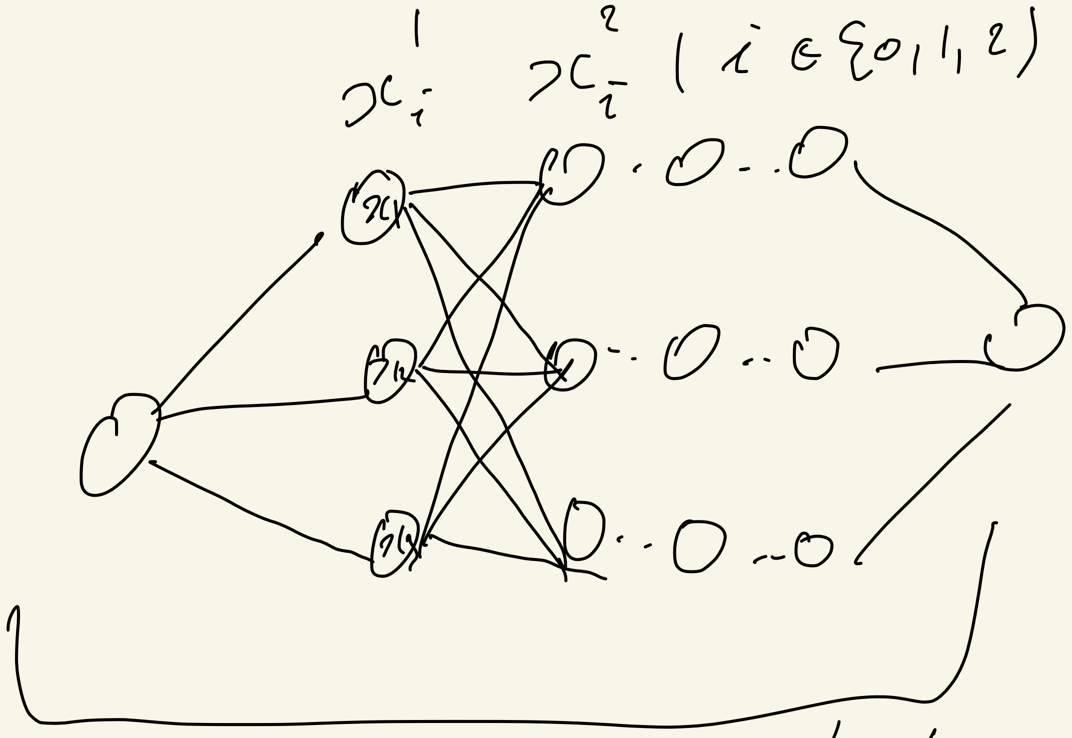
2. $L(y^{Pred}, y^{True})$

3. Update w_i

4. Repect

$\boxed{\text{start } 12:10}$

Forward Propogation / Neural Networks



Real architecture

$$g(\theta, x) = ax^2 + bx + c \mid 3 \text{ params}$$

$$f(\theta, x) : i \times h_1 + \sum_{k}^{n-1} (h_k \times h_{k+1}) + h_n + O$$

$$+ \sum_{k}^{n} h_k + O$$

$$\begin{bmatrix} W_{00} & W_{01} & W_{02} \\ W_{10} & W_{11} & W_{12} \\ W_{20} & W_{21} & W_{22} \end{bmatrix} \begin{bmatrix} x_0^1 \\ x_1^1 \\ x_2^1 \end{bmatrix} = f\left(\begin{bmatrix} W_{0i} \cdot x_i^1 \\ W_{1i} \cdot x_i^1 \\ W_{2i} \cdot x_i^1 \end{bmatrix} \right)$$

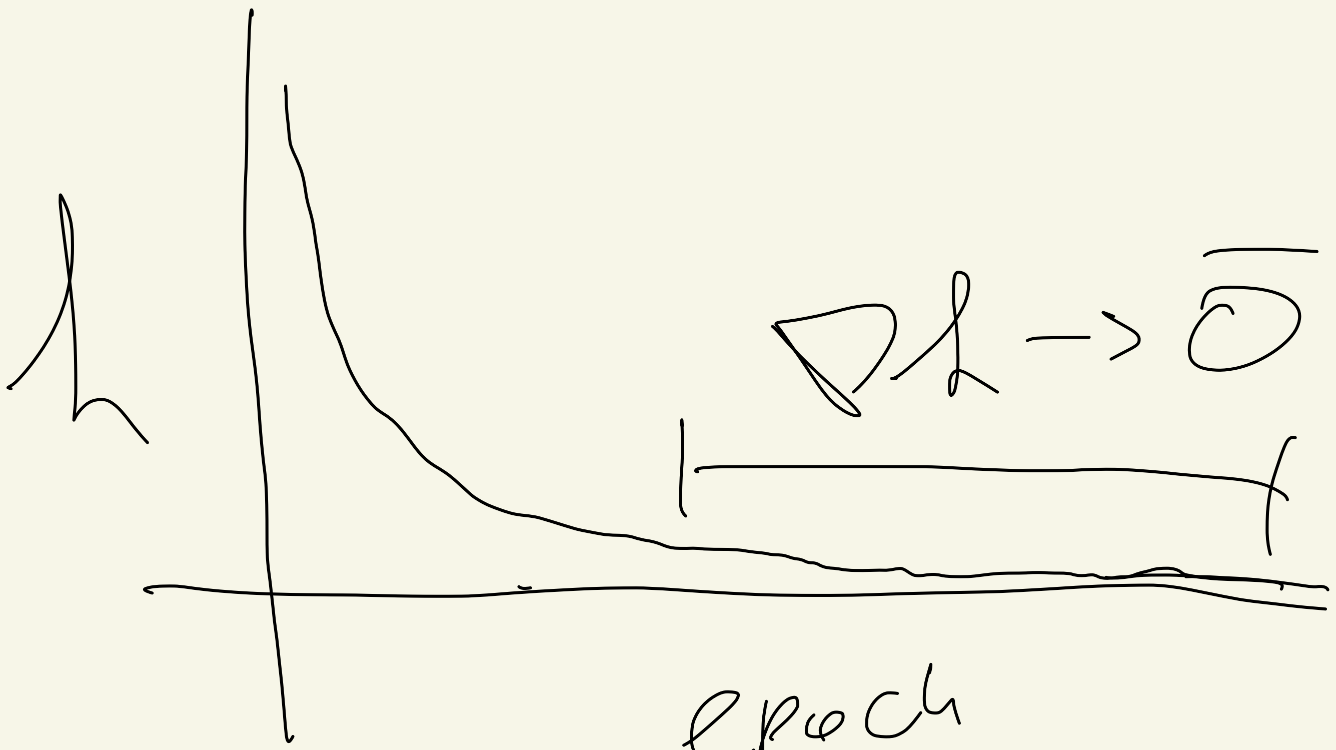
$\hookrightarrow$ Activation

Loss Functions

$$f(\{\theta\}, \bar{x}) = y_{\{\theta\}} \quad \Big| \quad y^{true}$$

$$L\left(f(\{\theta\}, \bar{x}), y^{true}\right) = \| y^{true} - y_{\{\theta\}} \|$$

$$\bar{x} = \bar{x}^{t} + \sigma_{noise}$$

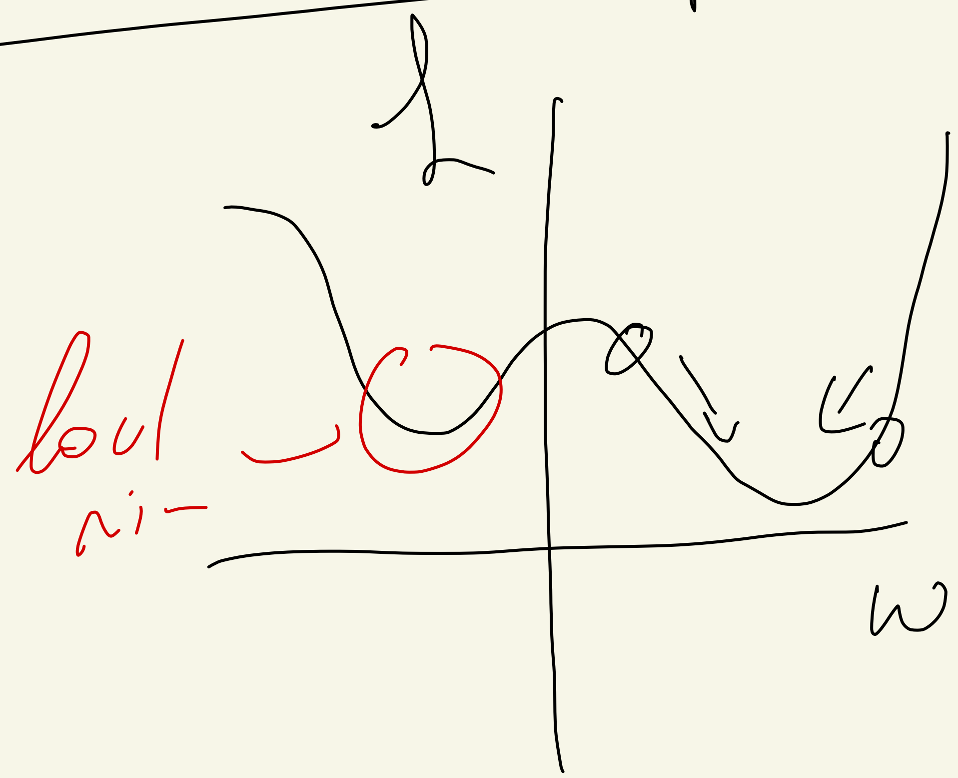


Batching => Issues with Sig.

$$L = \frac{1}{N} \sum_{i} \left\{ L[f(\{\theta\}, x_i), y_i^{true}] \right\}$$

$$\therefore \langle \sigma \rangle = 0$$

Network updates



→ SGD

$$w^{new} = w^{old} - \eta \nabla_w L$$

↳ learning rate

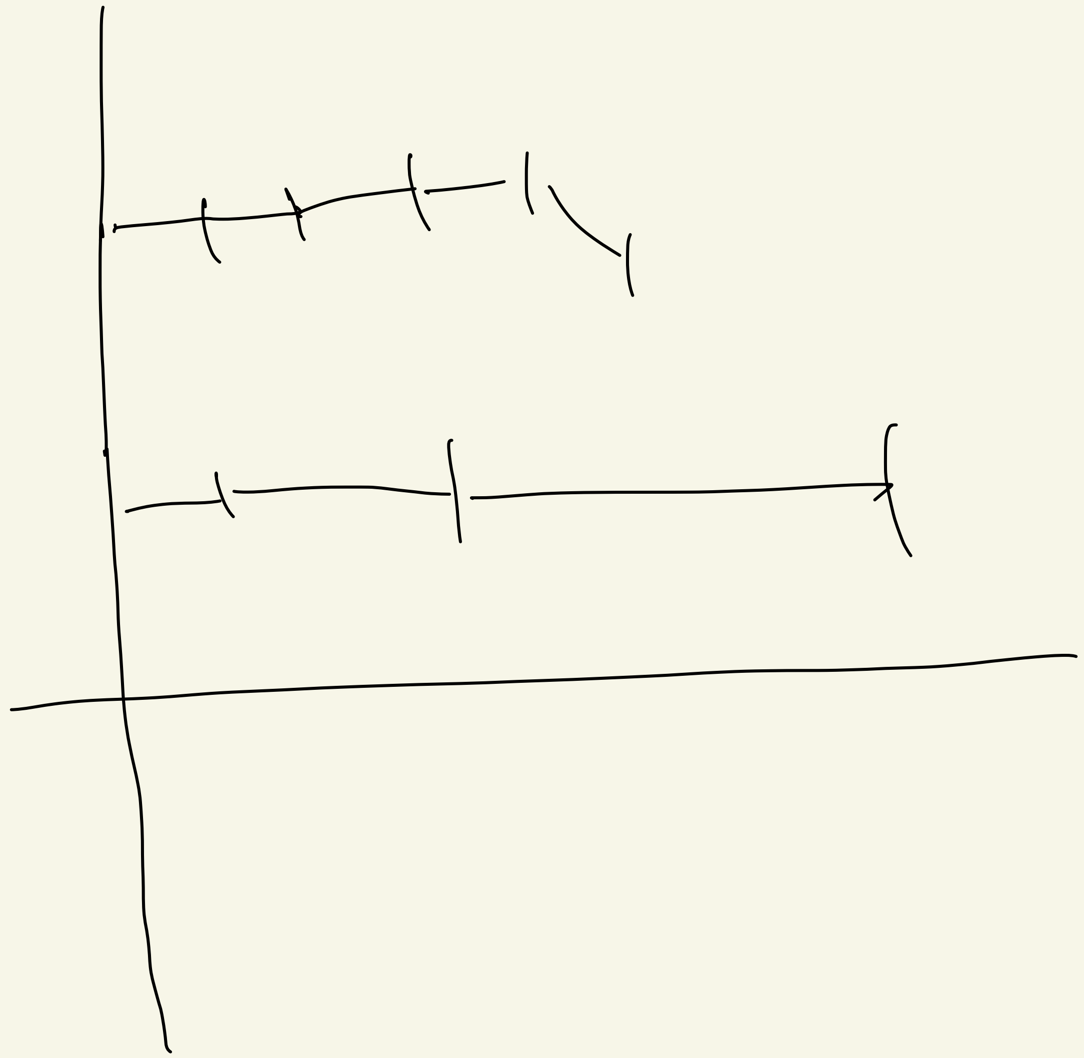
→ stuck in local minim

→ Slow

→ Momentum

$$\theta_{t+1} = \theta_t - \eta \nabla L + \gamma \sum_{i}^{t} v_t$$

$$v_t = \eta \nabla L_{\theta_t}$$



Adam Optimizer (Adaptive Moment Est.)

"start with large step, finish with small ones"

$$\theta_{t+1} = \theta_t - \eta \frac{\hat{m}_{t+1}}{(\hat{v}_{t+1}^{1/2} + \varepsilon)}$$

$$m_{t+1} = \beta_1 m_t + (1 - \beta_1) \cdot \nabla_{\theta_t} L$$

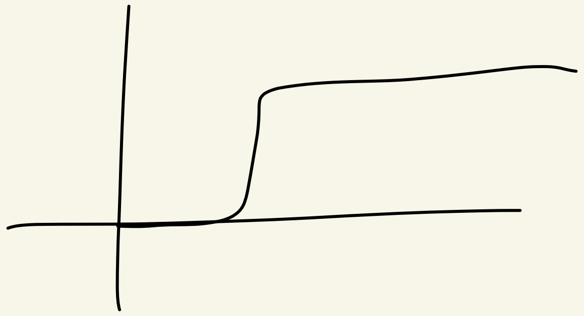
$$v_{t+1} = \beta_2 v_t + (1 - \beta_2) \cdot \nabla_{\theta_t} L$$

$$\hat{m}_{t+1} = \frac{m_t}{1 - \beta_1^t} \; ; \; \hat{v}_t = \frac{v_t}{1 - \beta_2^t}$$

$$\boxed{m_0 = v_0 = 0} \longrightarrow \text{Initial Conditions}$$

Activation Functions

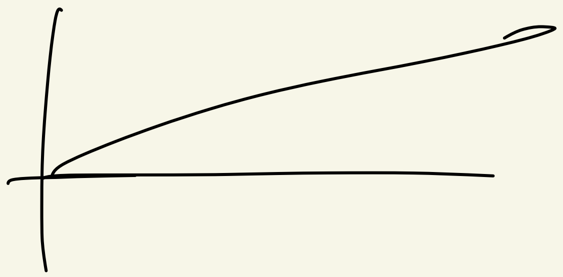
→ Heaviside



→ not-differentiable

→ too simple

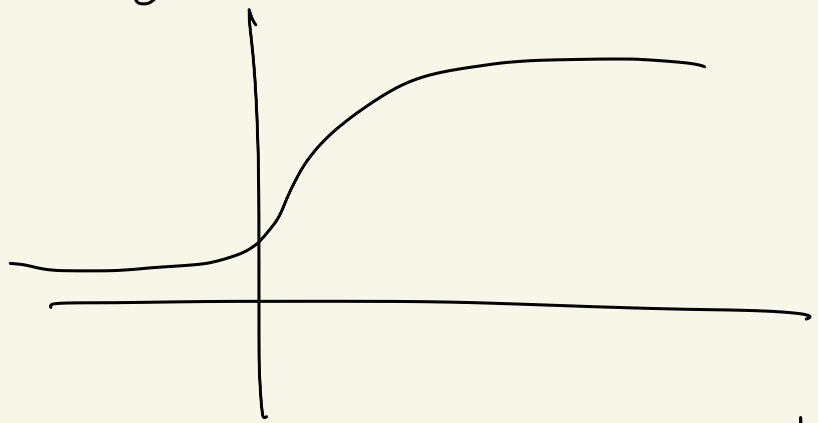
→ linear



→ No ∂x dep

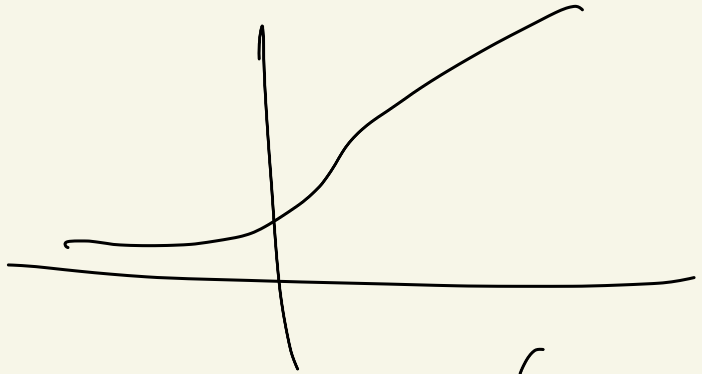
$\partial y / \partial x = 6$

→ Sigmoid



→ vanishing grad

→ ReLU



→ (0, Inf)

. . .

Auto-diff

→ Numeric Differentiation

$$\frac{\partial f(x)}{\partial x_i} \sim \frac{f(x+h) - f(x)}{h}$$

→ Symbolic Differentiation

$$\frac{d}{dx}(f(x) + g(x)) = \frac{df}{dx} + \frac{dg}{dx}$$

$$\frac{dx^2}{dx} = 2x$$

Auto-Diff 2

$$f(x_1, x_2) = \boxed{\ln(x_1) + x_1 x_2 - \sin(x_2)}$$

$g(f(x_1, x_2))$

$$\underbrace{v_0 = \ln(x_1)}, \quad \underbrace{\dot{v}_0 = \frac{1}{x_1}}$$

$$\underbrace{v_1 = x_1 x_2}, \quad \underbrace{\dot{v}_1\ x_2}$$

$$\underbrace{v_2 = -\sin(x_2)}, \quad \underbrace{\dot{v}_2 = 0}$$

$$\Rightarrow \frac{\partial f}{\partial x_1} = \dot{v}_0 + \dot{v}_1 + \dot{v}_2$$

$$= \frac{1}{x_1} + x_2$$

Dual Numbers

$\forall\, x_i$ let $x_i \to \underline{x_i} + \dot{x}_i \underline{\varepsilon}$: $\varepsilon^2 = 0$

$X = x + \dot{x}\varepsilon$, $Y = y + \dot{y}\varepsilon$

$$X \cdot Y = (x + \dot{x}\varepsilon)(y + \dot{y}\varepsilon)$$

$$= \underbrace{xy + \dot{y}x\varepsilon + \dot{x}y\varepsilon}_{} + \cancel{\dot{x}\dot{y}\varepsilon^2}^{=0}$$

$\hookrightarrow$ Product rule for $\dot{x} = \dfrac{\partial x}{\partial \varepsilon}$

Physics Aware ML (NN)

Data

$x \in \mathbb{R}^n$

→ Norm / Stand

→ feat. extr

→ Descriptors

$f(x) = f(\hat{P}x)$

Model

$f(x)$

→ layer

→ activation

→ Model $Af(x)$

→ layer-wise

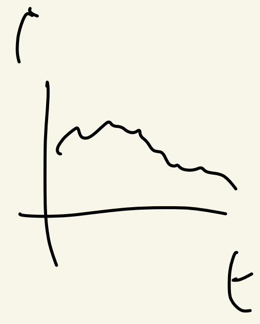
→ $f(\hat{P}x) = \hat{P}' f(x)$

Loss

Loss

$L \propto f(x) - f(x)$

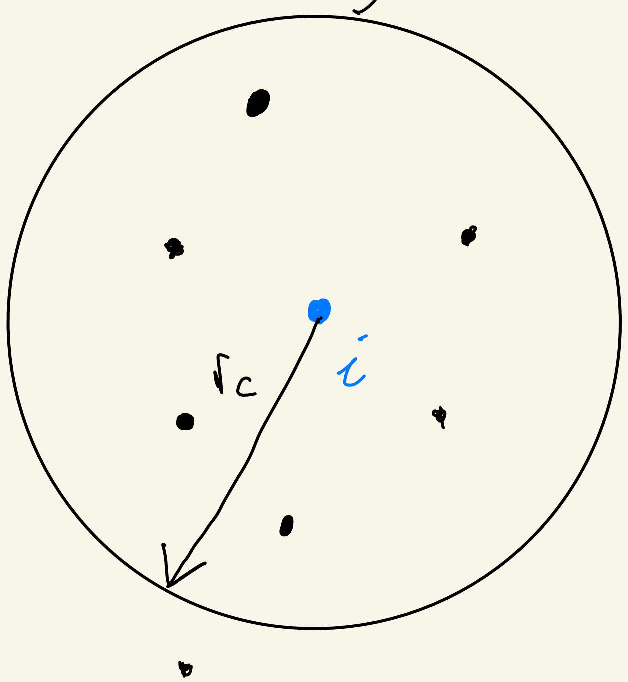
Deploy



Descriptors of Atomic Environments

$$E = \sum \varepsilon_i$$

↳ Atomic energy



$$\alpha_i = (\bar{r}_i, z_i, \bar{v}_i, \eta_i, \dots)$$

$$U = \{\alpha_0, \alpha_1, \dots, \alpha_N\}$$

$$A_{\alpha_i} = \{\alpha_j : d(\alpha_i, \alpha_j) \leq r_c \vee i \neq j\}$$

↳ Atomic environment of atom i

$$A = \{A_{\alpha_i} \ \forall \ \alpha_i \in U\} \therefore f \ A_{\alpha_i} \to \varepsilon_{\alpha_i}$$

Descriptors of atomic environments II

$$\hat{R}_\theta A_{\alpha_i} \to A_{\alpha_i}\,, \quad \hat{T}_{\bar{x}} A_{\alpha_i} \to A_{\alpha_i}$$

$$\mathcal{T} = \{ \hat{R}_\theta, \hat{T}_{\bar{x}}, \hat{P}_{\alpha\alpha}, \ldots \}$$

$$g : \mathcal{A} \to \mathcal{D} \;;\; \mathcal{D} = \{ d_i : \hat{A} d_i \to d_i \;\forall \hat{A} \in \mathcal{T} \}$$

↳ Descriptor space

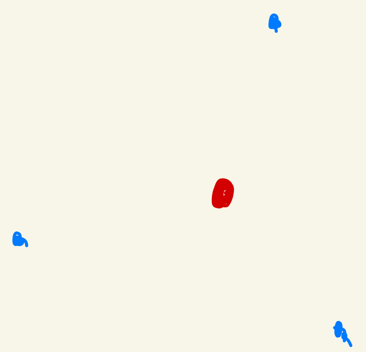
Descriptors of Atomic Environments III

Conditions

1. Describe an atomic environment
2. Fixed length
3. Sensitive to relevant changes
4. Respect relevant invariances*

↳ Physics aware ML

Smooth Overlap of Atomic Positions



$$\rho(\bar{r}) = \sum_{nlm} c_{nlm}\, g_n(r)\, Y_{lm}(\hat{r})$$

$$P_{nn'l} = \sum_{m} c_{nlm} \left(c_{n'lm}\right)^{*}$$

↳ Power spectrum

$$k\left(P_{nn'l}, P'_{nn'l}\right) = \left(\frac{P_{nn'l} \cdot P'_{nn'l}}{\|P_{nn'l}\| \cdot \|P'_{nn'l}\|}\right)^{\xi}$$

ξ → Sensitivity index