DiLoc
?
$\mathcal{D} =$
$(\mathbf{X}, \mathbf{y})$
$\mathbf{X}$
$\mathbf{y}$
$f(\mathbf{x};\theta)$
$f :$
$\mathbf{x} \to$
y
θ
$\mathbf{y}$
$\mathcal{C}(\mathbf{y}, f(\mathbf{X});\theta)$
non-lin-ear ma-chine learn-ing mod-els
?
?
??
$_architecture.png$ **(A)** $The fundamental structure of neural networks comprises simplified neuron units that perform a linea$
$linear activation function.$ **(B)** $These neuron units are organized into layers, where the output of one layer serves as the input to$

$$a = f\left(\Sigma_{i=1}^{n} w_i x_i\right)\right) = f(z) \tag{1}$$

a
f
$x_i, x_{i+1}, ..., x_n$
n
$w_i, w_{i+1}, ..., w_n$
$b_i, b_{i+1}, ..., b_n$
a
a
$f(z)$
??
z^i

$$z^i = w^{(i)} \cdot x + b^{(i)} = \mathbf{x}^T \cdot \mathbf{w}^{(i)}, \tag{2}$$

$\mathbf{x} =$
$(1, x)$
$\mathbf{w}^i =$
$(b^{(i)}), w^{(i)})$
f_i

$$a_i(\mathbf{x}) = f_i(z^{(i)}). \tag{3}$$

??
$_architecture.pdf Architecture.$
$\mathcal{D}$
%
?
θ

$$MSE(\theta) = \frac{1}{n} \sum_{i=0}^{n-1} (y_i - \tilde{y}_i)^2, \tag{4}$$

$\theta =$
$\theta_0, \theta_1, ..., \theta_n$
$\tilde{y}_i$
y_i

$$C(\theta) = \Sigma_{i=0}^{n} c_i(\mathbf{x}_i, \theta) \tag{5}$$

$c_i(\mathbf{x}_i, \theta)$
i
$\mathbf{x}_i$
θ
n
cost func-tions
loss

$$MSE(\theta) = \frac{1}{N-1} \sum_{i=0}^{N} (y_i - \tilde{y}_i)^2, \tag{6}$$

$\theta =$
$\theta_0, \theta_1, ..., \theta_n$
y_i
y_i
?
?
?
??
μ
$_visualization_max.pdf$

$$MAE_r(\theta) = \frac{1}{n}\sum_{i=0}^{n-1} \|\theta_{r,i} - \tilde{\theta}_{r,i}\|, \tag{9}$$

m

$$MED_{x_1,y_1,z_1,\ldots,x_m,y_m,z_m}(\theta) = \left(\frac{1}{n}\sum_{i=0}^{n-1}\sqrt{(\theta_{x_1,i}-\tilde{\theta}_{x_1,i})^2+(\theta_{y_1,i}-\tilde{\theta}_{y_1,i})^2+(\theta_{z_1,i}-\tilde{\theta}_{z_1,i})^2}\right)+\left(\frac{1}{n}\sum_{i=0}^{n-1}\sqrt{(\theta_{x_2,i}-\tilde{\theta}_{x_2,i})^2+(\tag{10}$$

$$C(\theta) = \{\ arraylMED_{x,y,z}(\theta), if\|\theta\| = 3MED_{x,y,z}(\theta) + MAE_A(\theta), if\|\theta\| = 4MED_{x,y,z}(\theta) + MAE_A(\theta) + MAE_r(\theta), if\| \tag{11}$$

$|\theta|$
$|\theta| = 3$
$|\theta| = 4$
$|\theta| = 5$
$|\theta|$
%
%
θ
$F(\mathbf{x})$
$\mathbf{w}$
$-\nabla F(\mathbf{a})$

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \eta\nabla F(\mathbf{w}_n) \tag{12}$$

η
$F((w)_n) \geq F((w)_{n+1})$
?
$\mathbf{w}$
?
x_0
F
$\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$

$$\mathbf{x}_{n+1} = \mathbf{x}_n - \eta_n\nabla F(\mathbf{x}_n), n \geq 0, \tag{13}$$

$\eta_n \geq 0$

$$F(\mathbf{x}_0) \geq F(\mathbf{x}_1) \geq F(\mathbf{x}_2) \geq \ldots \geq F(\mathbf{x}_n) \tag{14}$$

$(\mathbf{x}_n)$
?
batch
$\mathbf{w}_0$
η
?
?

$$\mathbf{w}_{\tau+1} = \mathbf{w}_\tau - \eta\nabla F_n(\mathbf{w}_\tau) \tag{15}$$

mo-men-tum
η
γ

$$\mathbf{v}_\tau = \gamma\mathbf{v}_{\tau-1} - \eta\nabla F_n(\mathbf{w}_\tau) \tag{16}$$

$$\mathbf{w}_\tau = \mathbf{w}_{\tau-1} + \mathbf{v}_\tau \tag{17}$$

$\mathbf{w}\tau$
τ
$\mathbf{w}_{\tau-1}$
$\mathbf{v}_\tau$
τ
γ
η
$\nabla F_n(\mathbf{w}_\tau)$
F_n
β
?
?
?

$$f(x) = \frac{1}{1+e^{-x}} \tag{18}$$