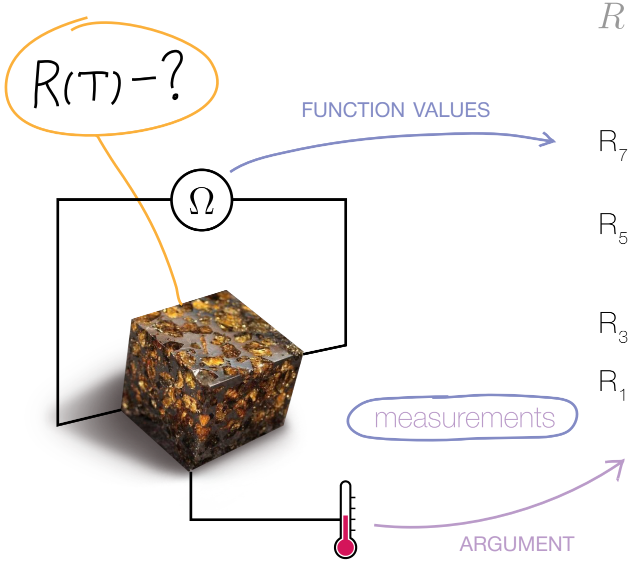
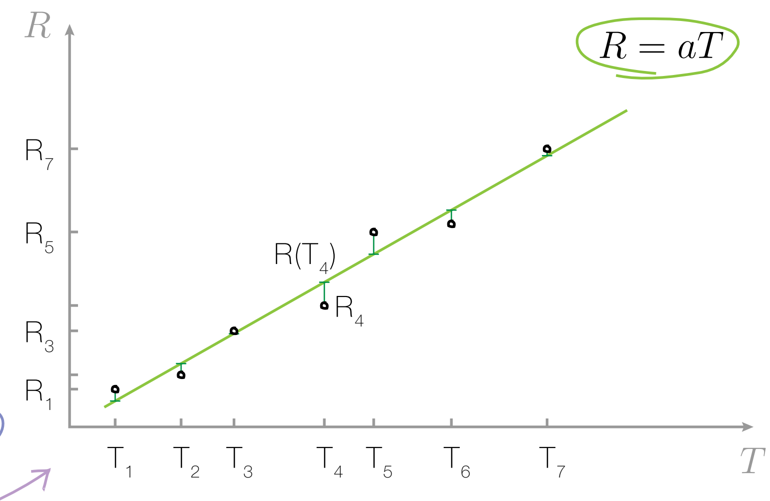
Approximating unknown functions

Physical properties

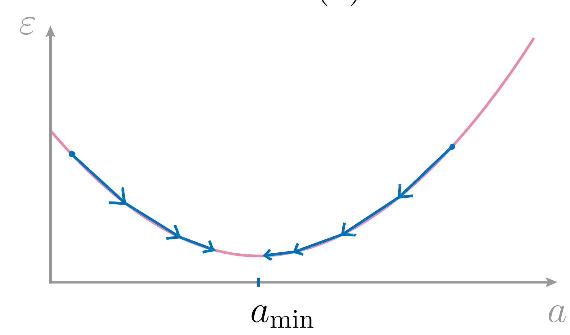




$$\varepsilon = \frac{1}{2}\sum_n \left(R\left(T_n\right) - R_n\right)^2 \qquad a:\ \varepsilon \xrightarrow[a]{} \min$$

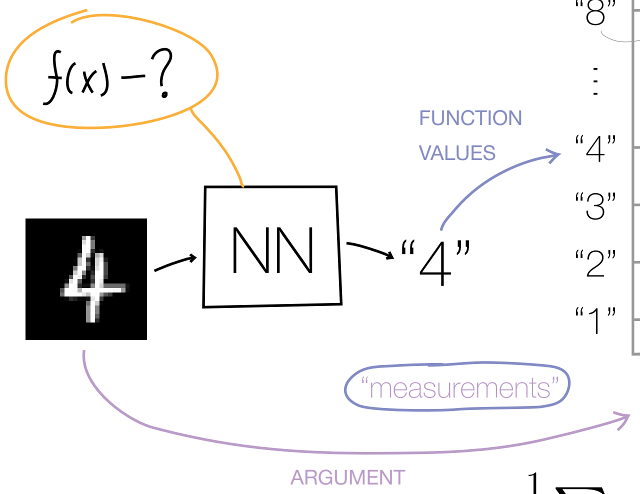
Gradient descent

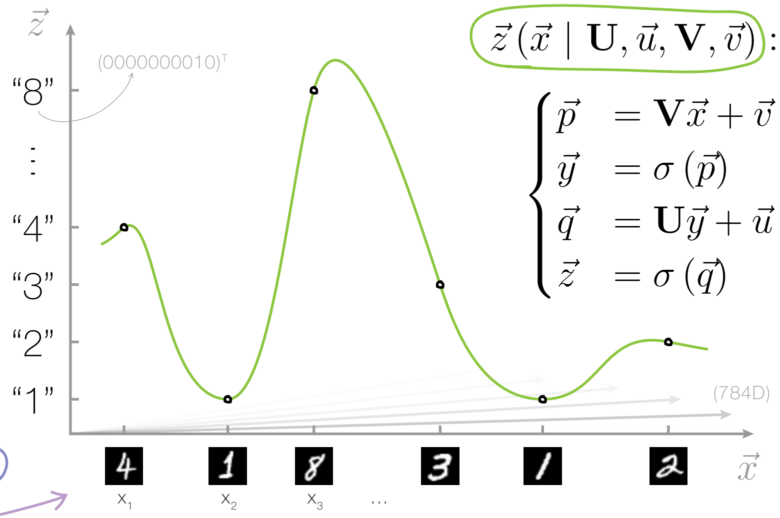
$$\varepsilon = \varepsilon\left(a\right)$$



$$a^{t+1} = a^t - \alpha \frac{d\varepsilon}{da}\left(a^t\right) \rightarrow a \xrightarrow[t\to+\infty]{} a_{\min}$$

Image recognition



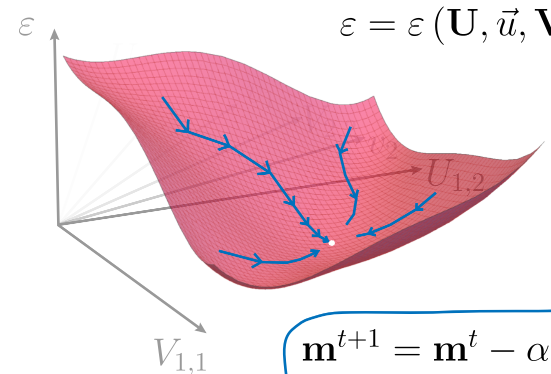


$$\vec{z}\left(\vec{x} \mid \mathbf{U}, \vec{u}, \mathbf{V}, \vec{v}\right):$$

$$\begin{cases} \vec{p} &= \mathbf{V}\vec{x} + \vec{v} \\ \vec{y} &= \sigma\left(\vec{p}\right) \\ \vec{q} &= \mathbf{U}\vec{y} + \vec{u} \\ \vec{z} &= \sigma\left(\vec{q}\right) \end{cases}$$

$$\varepsilon = \frac{1}{2}\sum_n \left(\vec{z}\left(\vec{x}_n\right) - \vec{z}_n\right)^2 \qquad \mathbf{U}, \vec{u}, \mathbf{V}, \vec{v}:\ \varepsilon \xrightarrow[\mathbf{U},\vec{u},\mathbf{V},\vec{v}]{} \min$$

$$\varepsilon = \varepsilon\left(\mathbf{U}, \vec{u}, \mathbf{V}, \vec{v}\right)$$



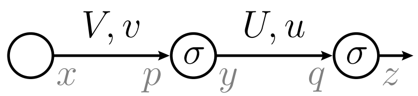
$$\mathbf{m}^{t+1} = \mathbf{m}^t - \alpha \frac{\partial\varepsilon}{\partial\mathbf{m}}$$

$$\mathbf{m} = \{\mathbf{U}, \vec{u}, \mathbf{V}, \vec{v}\} = \left\{\begin{matrix} U_{1,1}, U_{1,2}, \ldots, U_{2,1}, U_{2,2}, \ldots, u_1, u_2, \ldots, \\ V_{1,1}, V_{1,2}, \ldots, V_{2,1}, V_{2,2}, \ldots, v_1, v_2, \ldots \end{matrix}\right\}$$

[2] Stochastic gradient descent

Neural Network Training

Simplified one-dimensional case



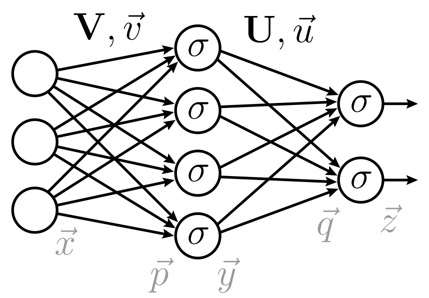
$$
\begin{aligned}
p &= Vx + v \\
y &= \sigma(p) \\
q &= Uy + u \\
z &= \sigma(q) \\
\varepsilon &\overset{(\Sigma)}{=} \frac{1}{2}(z - l)^2
\end{aligned}
$$

$$
\frac{\partial \varepsilon}{\partial \mathbf{m}}
\begin{cases}
\dfrac{\partial \varepsilon}{\partial V} = \dfrac{\partial \varepsilon}{\partial p}\dfrac{\partial p}{\partial V} = \dfrac{\partial \varepsilon}{\partial p} x \\[2ex]
\dfrac{\partial \varepsilon}{\partial v} = \dfrac{\partial \varepsilon}{\partial p}\dfrac{\partial p}{\partial v} = \dfrac{\partial \varepsilon}{\partial p} = \dfrac{\partial \varepsilon}{\partial y}\dfrac{\partial y}{\partial p} = \dfrac{\partial \varepsilon}{\partial y}\sigma'(p) \\[2ex]
\dfrac{\partial \varepsilon}{\partial y} = \dfrac{\partial \varepsilon}{\partial q}\dfrac{\partial q}{\partial y} = \dfrac{\partial \varepsilon}{\partial q} U \\[2ex]
\dfrac{\partial \varepsilon}{\partial U} = \dfrac{\partial \varepsilon}{\partial q}\dfrac{\partial q}{\partial U} = \dfrac{\partial \varepsilon}{\partial q} y \\[2ex]
\dfrac{\partial \varepsilon}{\partial u} = \dfrac{\partial \varepsilon}{\partial q}\dfrac{\partial q}{\partial u} = \dfrac{\partial \varepsilon}{\partial q} = \dfrac{\partial \varepsilon}{\partial z}\dfrac{\partial z}{\partial q} = \dfrac{\partial \varepsilon}{\partial z}\sigma'(q)
\end{cases}
$$

$$\frac{\partial \varepsilon}{\partial z} = (z - l)$$

Back Propagation

Full derivation



$$
\begin{aligned}
\vec{p} &= \mathbf{V}\vec{x} + \vec{v} \\
\vec{y} &= \sigma(\vec{p}) \\
\vec{q} &= \mathbf{U}\vec{y} + \vec{u} \\
\vec{z} &= \sigma(\vec{q}) \\
\varepsilon &\overset{(\Sigma)}{=} \frac{1}{2}\left(\vec{z} - \vec{l}\right)^2
\end{aligned}
$$

$$
\frac{\partial \varepsilon}{\partial \mathbf{m}}
\begin{cases}
\dfrac{\partial \varepsilon}{\partial \mathbf{V}} = \dfrac{\partial \varepsilon}{\partial V_{ij}} = \sum_k \dfrac{\partial \varepsilon}{\partial p_k}\dfrac{\partial p_k}{\partial V_{ij}} = \dfrac{\partial \varepsilon}{\partial p_i}\dfrac{\partial p_i}{\partial V_{ij}} = \dfrac{\partial \varepsilon}{\partial p_i} x_j = \dfrac{\partial \varepsilon}{\partial \vec{p}} \otimes \vec{x} \\[2ex]
\dfrac{\partial \varepsilon}{\partial \vec{v}} = \dfrac{\partial \varepsilon}{\partial v_i} = \dfrac{\partial \varepsilon}{\partial p_i}\dfrac{\partial p_i}{\partial v_i} = \dfrac{\partial \varepsilon}{\partial p_i} = \dfrac{\partial \varepsilon}{\partial y_i}\dfrac{\partial y_i}{\partial p_i} = \dfrac{\partial \varepsilon}{\partial y_i}\sigma'(p_i) = \dfrac{\partial \varepsilon}{\partial \vec{y}} \odot \sigma'(\vec{p}) \\[2ex]
\dfrac{\partial \varepsilon}{\partial \vec{y}} = \dfrac{\partial \varepsilon}{\partial y_i} = \sum_j \dfrac{\partial \varepsilon}{\partial q_j}\dfrac{\partial q_j}{\partial y_i} = \sum_j \dfrac{\partial \varepsilon}{\partial q_j} U_{ji} = \sum_j U^T_{ij}\dfrac{\partial \varepsilon}{\partial q_j} = \mathbf{U}^T \dfrac{\partial \varepsilon}{\partial \vec{q}} \\[2ex]
\dfrac{\partial \varepsilon}{\partial \mathbf{U}} = \dots = \dfrac{\partial \varepsilon}{\partial \vec{q}} \otimes \vec{y} \\[2ex]
\dfrac{\partial \varepsilon}{\partial \vec{u}} = \dots = \dfrac{\partial \varepsilon}{\partial q_i} = \dfrac{\partial \varepsilon}{\partial \vec{q}} = \dots = \dfrac{\partial \varepsilon}{\partial \vec{z}} \odot \sigma'(\vec{q})
\end{cases}
$$

$$\frac{\partial \varepsilon}{\partial \vec{z}} = \vec{z} - \vec{l}$$

column × row = matix (Tensor product)

element-wise (Hadamard product)

$q_j = \sum U_{ji} y_i$

matrix × column = column

Back Propagation

$x/\vec{x}$ — image vector (ARGUMENT)		Activation function	Derivative
$l/\vec{l}\ (= z_n/\vec{z}_n)$ — label (VALUE)		$\sigma(x) = \frac{1}{1+e^{-x}}$	$\sigma'(x) = \frac{e^{-x}}{(1+e^{-x})^2}$