Pushforward / Jvp rule for the L_2-Loss

↳ squared L_2 norm

$$\ell(\underline{y}, \underline{y}^r) = \frac{1}{2} \| \underline{y} - \underline{y}^r \|_2^2 =: \mathcal{L}$$

$\underline{y} \in \mathbb{R}^N$ ↳ typical in regression

$\underline{y}^r \in \mathbb{R}^N$

$\mathcal{L} \in \mathbb{R}$

$\ell : \mathbb{R}^N \times \mathbb{R}^N \rightarrow \mathbb{R}$

task: forward propagate tangents on the input $\underline{\dot{y}} \in \mathbb{R}^N$ & $\underline{\dot{y}}^r \in \mathbb{R}^N$ to the output $\dot{\mathcal{L}} \in \mathbb{R}$

in general

$$\dot{\mathcal{L}} = \frac{\partial \ell}{\partial \underline{y}} \underline{\dot{y}} + \frac{\partial \ell}{\partial \underline{y}^r} \underline{\dot{y}}^r$$

$$\frac{\partial \ell}{\partial \underline{y}} = \frac{\partial}{\partial \underline{y}} \left(\frac{1}{2} \| \underline{y} - \underline{y}^r \|_2^2 \right) = \frac{\partial}{\partial \underline{y}} \left(\frac{1}{2} (\underline{y} - \underline{y}^r)^T (\underline{y} - \underline{y}^r) \right)$$

index notation

$$\frac{\partial \ell}{\partial y_i} = \frac{\partial}{\partial y_i} \left(\frac{1}{2} \sum_{j=1}^{N} (y_j - y_j^r)^2 \right)$$

$$= \frac{1}{2} \sum_{j=1}^{N} \frac{\partial}{\partial y_i} (y_j - y_j^r)^2$$

$$= \frac{1}{2} \sum_{j=1}^{N} 2 (y_j - y_j^r) \frac{\partial y_j}{\partial y_i} \qquad = \delta_{ji} \qquad \left(\frac{\partial - y_j^r}{\partial y_i} \right)$$

$$= \frac{1}{2} \sum_{j=1}^{N} 2 (y_j - y_j^r) \delta_{ji}$$

$$= \frac{1}{2} 2 (y_i - y_i^r)$$

$$= y_i - y_i^r$$

back to symbolic notation

$$\frac{\partial \ell}{\partial \underline{y}} = (\underline{y} - \underline{y}^r)^T \qquad \frac{\partial \ell}{\partial \underline{y}^r} = (\underline{y}^r - \underline{y})^T = -(\underline{y} - \underline{y}^r)^T$$

$$\dot{\mathcal{L}} = (\underline{y} - \underline{y}^r)^T \underline{\dot{y}} + (\underline{y}^r - \underline{y})^T \underline{\dot{y}}^r$$

Full pushforward rule

$$\boxed{\mathfrak{F}(\ell, (\underline{y}, \underline{y}^r), (\underline{\dot{y}}, \underline{\dot{y}}^r)) = \left(\underbrace{\left(\frac{1}{2} \| \underline{y} - \underline{y}^r \|_2^2, \right)}_{\mathcal{L}}, \underbrace{\left((\underline{y} - \underline{y}^r)^T \underline{\dot{y}} + (\underline{y}^r - \underline{y})^T \underline{\dot{y}}^r \right)}_{\dot{\mathcal{L}}} \right)}$$

often this $\underline{\dot{y}}^r$ is $\underline{0}$

$$\Delta \underline{y} = \underline{y} - \underline{y}^r$$