Pushforward / Jvp rule for softmax

$$f(\underline{x}) = \underline{y} \qquad y_i = \frac{e^{x_i}}{\sum_j' e^{x_j}}$$

$$\underline{x} \in \mathbb{R}^N$$

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

task: forward propagate $\dot{\underline{x}} \in \mathbb{R}^N$ to $\dot{\underline{y}} \in \mathbb{R}^N$

$$\dot{\underline{y}} = \frac{\partial \underline{y}}{\partial \underline{x}} \dot{\underline{x}}$$

$$\dot{y}_i = \frac{\partial y_i}{\partial x_k} \dot{x}_k$$

$$\frac{\partial y_i}{\partial x_k} = \frac{\partial}{\partial x_k}\left(\frac{e^{x_i}}{\sum_j' e^{x_j}} \right)$$

$$\left(\frac{u}{v}\right)' = \frac{u'v - uv'}{v^2}$$

$$= e^{x_k}$$

$$\frac{\partial y_i}{\partial x_k} = \frac{e^{x_i}\delta_{ik} \sum_j' e^{x_j} - e^{x_i} \sum_j' e^{x_j}\delta_{jk}}{\left(\sum_j' e^{x_j}\right)^2}$$

$$= \frac{e^{x_i}\delta_{ik} \cancel{\sum_j' e^{x_j}}}{\left(\sum_j' e^{x_j}\right)^{\cancel{2}}} - \frac{e^{x_i} \, e^{x_k}}{\left(\sum_j' e^{x_j}\right)^2}$$

$$= y_i \delta_{ik} - y_i y_k$$

-> primal output "y" appears again in its derivative

$$\Rightarrow \dot{y}_i = (y_i \delta_{ik} - y_i y_k)\dot{x}_k$$

$$= y_i \delta_{ik} \dot{x}_k - y_i y_k \dot{x}_k$$

$$= y_i \dot{x}_i - y_i y_k \dot{x}_k$$

in symbolic notation

$$\dot{\underline{y}} = \underline{y} \odot \dot{\underline{x}} - (\underline{y}^T \dot{\underline{x}})\, \underline{y}$$

full pushforward rule

$$\boxed{\hat{f}\left(\text{softmax}, (\underline{x},), (\dot{\underline{x}},)\right) = \left(\big(\underbrace{\text{softmax}(\underline{x})}_{\underline{y}},\big), \big(\underbrace{\underline{y} \odot \dot{\underline{x}} - \underline{y}(\underline{y}^T \dot{\underline{x}})}_{\dot{\underline{y}}}\big) \right)}$$