(a) 2.

$$CE(y,\hat{y}) = -\sum_i y_i \log(\hat{y}_i)$$

$$\text{Softmax}(\theta)_i = \frac{\exp(\theta_i)}{\sum_j \exp(\theta_j)}$$

$$\frac{\partial CE(y,\hat{y})}{\partial \theta_i} = \frac{\partial}{\partial \theta_i}\left(-\sum_k y_k \log(\hat{y}_k)\right) =$$

$$\underset{\text{כלל השרשרת}}{=} \frac{\partial\left(-\sum_k y_k \log(\hat{y}_k)\right)}{\partial \log(\hat{y}_k)} \cdot \frac{\partial \log(\hat{y}_k)}{\partial \hat{y}_k} \cdot \frac{\partial \hat{y}_k}{\partial \theta_i} =$$

$$= -\sum_k \frac{\partial(y_k \log(\hat{y}_k))}{\partial \log(\hat{y}_k)} \cdot \frac{1}{\hat{y}_k} \cdot \frac{\partial \hat{y}_k}{\partial \theta_i} =$$

$$= -\sum_k y_k \cdot \frac{1}{\hat{y}_k} \cdot \frac{\partial \hat{y}_k}{\partial \theta_i} = \ldots$$

$$\frac{\partial \hat{y}_k}{\partial \theta_i} = \frac{\partial}{\partial \theta_i} \frac{e^{\theta_k}}{\sum_j e^{\theta_j}}$$

נפרק לשני מקרים:

• $k=i$:

$$\frac{\partial}{\partial \theta_i} \frac{e^{\theta_i}}{\sum_j e^{\theta_j}} \underset{\substack{\text{גם במונה} \\ \text{וגם במכנה} \\ \text{מופיע } \theta_i}}{=} \frac{e^{\theta_i} \sum_j e^{\theta_j} - e^{\theta_i} e^{\theta_i}}{\left(\sum_j e^{\theta_j}\right)^2} = \frac{e^{\theta_i}}{\sum_j e^{\theta_j}} \cdot \frac{\sum_j e^{\theta_j} - e^{\theta_i}}{\sum_j e^{\theta_j}} =$$

$$= \hat{y}_i \cdot (1-\hat{y}_i)$$

• $k \neq i$:

$$\frac{\partial}{\partial \theta_i} \frac{e^{\theta_k}}{\sum_j e^{\theta_j}} \underset{\substack{\text{רק במכנה} \\ \text{מופיע } \theta_i}}{=} \frac{-e^{\theta_k} \cdot e^{\theta_i}}{\left(\sum_j e^{\theta_j}\right)^2} = -\frac{e^{\theta_k}}{\sum_j e^{\theta_j}} \cdot \frac{e^{\theta_i}}{\sum_j e^{\theta_j}} = -\hat{y}_k \hat{y}_i$$

נאחד לתוצאה:

$$\frac{\partial}{\partial \theta_i} \frac{e^{\theta_k}}{\sum_j e^{\theta_j}} = \begin{cases} \hat{y}_i(1-\hat{y}_i), & k=i \\ -\hat{y}_k \hat{y}_i, & k \neq i \end{cases} = \hat{y}_k(\underbrace{\delta_{ki}}_{\text{פונ' דלתא קרונקר}} - \hat{y}_i)$$

$$\ldots = -\sum_k y_k \cdot \cancel{\frac{1}{\hat{y}_k}} \cdot \cancel{\hat{y}_k}(\delta_{ki} - \hat{y}_i) =$$

המשך בדף -

$$= -y_i \cdot (1-\hat{y}_i) + \sum_{k\neq i} y_k \hat{y}_i = -y_i + y_i \hat{y}_i + \sum_{k \neq i} y_k \hat{y}_i =$$

$$= -y_i + \hat{y}_i \left(y_i + \sum_{k\neq i} y_k \right) =$$

$$= \hat{y}_i \underbrace{\left(\sum_k y_k \right)}_{\text{one hot vector}} - y_i = \boxed{\hat{y}_i - y_i}$$

one hot vector

לכן הסכום שווה ל-1

$$\frac{\partial J}{\partial x} \underset{\text{כלל השרשרת}}{=} \frac{\partial\, CE(y,\hat{y})}{\partial (hW_2 + b_2)} \cdot \frac{\partial (hW_2 + b_2)}{\partial (\sigma(xW_1 + b_1))} \cdot \frac{\partial (\sigma(xW_1+b_1))}{\partial (xW_1 + b_1)} \cdot \quad (\text{ג.})$$

$$\cdot \frac{\partial (xW_1 + b_1)}{\partial x} \underset{a \text{ סעיף}}{=} (\hat{y} - y) \cdot \frac{\partial [\sigma(xW_1+b_1) W_2 + b_2]}{\partial (\sigma(xW_1+b_1))} \cdot \sigma'(xW_1+b_1) \cdot W_1^T =$$

$$= (\hat{y} - y) \cdot W_2^T \cdot \sigma'(xW_1 + b_1) \cdot W_1^T =$$

$$= \boxed{(\hat{y}-y) W_2^T \left[\sigma(xW_1+b_1) \cdot (1 - \sigma(xW_1+b_1)) \right] W_1^T.}$$