Theoretical background: Prove that softmax is invariant to constant offset in the input, i.e prove that for any input vector x and any constant c, $softmax(x) = softmax(x+c)$

Proof:

Let $softmax(x_i + c) = \frac{e^{x_i+c}}{\sum_{j=1}^{N} e^{x_j+c}}$

$\frac{e^{x_i+c}}{\sum_{j=1}^{N} e^{x_j+c}} = \frac{e^{x_i}e^c}{\sum_{j=1}^{N} e^{x_j}e^c}$ factoring out e^c

$\frac{e^{x_i}e^c}{\sum_{j=1}^{N} e^{x_j}e^c} = \frac{e^{x_i}e^c}{e^c\sum_{j=1}^{N} e^{x_j}}$

$\frac{e^{x_i}e^c}{e^c\sum_{j=1}^{N} e^{x_j}} = \frac{e^{x_i}}{\sum_{j=1}^{N} e^{x_j}}$ canceling e^c in numerator and denominator

$\frac{e^{x_i}}{\sum_{j=1}^{N} e^{x_j}} = softmax(x_i)$ By definition