2.) Given the softmax function in (10.13) and (4.13):

(a.) If in (10.13): $f_m(X) = Pr(Y=m \mid X) = e^{Z_m} / \sum_{l=0}^{9} e^{Z_l}$ $\circledast$

And we add a constant c to each Z_l, we get:

$$f_m^*(X) = \frac{e^{Z_m+c}}{\sum_{l=0}^{9} e^{Z_l+c}} = \frac{e^{Z_m+c}}{\sum_{l=0}^{9} e^{Z_l} e^{c}} = \frac{e^{Z_m} \cdot \cancel{e^{c}}}{\left(\sum_{l=0}^{9} e^{Z_l}\right) \cdot \cancel{e^{c}}} = \frac{e^{Z_m}}{\sum_{l=0}^{9} e^{Z_l}}$$

which is the same as the above $\circledast$ before adding c to each Z_l.

(b) If in (4.13): $$Pr(Y=k \mid X=x) = \frac{e^{\beta_{k0}+\beta_{k1}X_1+\dots+\beta_{kp}X_p}}{\sum_{l=1}^{K} e^{\beta_{l0}+\beta_{l1}X_1+\dots+\beta_{lp}X_p}} = f$$

Then adding constants c_j, $j=0,1,\dots p$ to the coefficients for each of the classes, then:

$$f^* = Pr(Y=k \mid X=x) = \frac{e^{(\beta_{k0}+c_0)+(\beta_{k1}+c_1)X_1+\dots+(\beta_{kp}+c_p)X_p}}{\sum_{l=1}^{K} e^{(\beta_{l0}+c_0)+(\beta_{l1}+c_1)X_1+\dots+(\beta_{lp}+c_p)X_p}}$$

$$= \frac{e^{(\beta_{k0}+\beta_{k1}X_1+\dots \beta_{kp}X_p)+(c_0+c_1X_1+\dots+c_pX_p)}}{\sum_{l=1}^{K} e^{(\beta_{l0}+\beta_{l1}X_1+\dots+\beta_{lp}X_p)+(c_0+c_1X_1+\dots+c_pX_p)}}$$

$$= \frac{\cancel{e^{c_0+c_1X_1+\dots+c_pX_p}} \cdot e^{\beta_{k0}+\beta_{k1}X_1+\dots+\beta_{kp}X_p}}{\cancel{e^{c_0+c_1X_1+\dots+c_pX_p}} \cdot \sum_{l=1}^{K} e^{\beta_{l0}+\beta_{l1}X_1+\dots+\beta_{lp}X_p}} = \boxed{\frac{e^{\beta_{k0}+\beta_{k1}X_1+\dots+\beta_{kp}X_p}}{\sum_{l=1}^{K} e^{\beta_{l0}+\beta_{l1}X_1+\dots+\beta_{lp}X_p}}}$$

↓ which is the same as above f.

Thus, when we add constants c_j, predictions at any new point x are unchanged.