

b) Derive $\frac{\partial}{\partial \theta_i} \ell(\theta; D)$

$$\frac{\partial}{\partial \theta_i} \ell(\theta; D) = \frac{1}{|D|} \sum_{(x,y) \in D} f_i(x,y) - (1-a) \cdot \left[\frac{1}{|D|} \sum_{x \in D} E_a[f_i(x,y)] \right] \\ - a \cdot \left[\frac{1}{|D|} \sum_{y \in D} E_a[f_i(x,y)] \right]$$

If D is the empirical distribution of our dataset \hat{P} ,
or a conditional distribution of the form $P_a(W|Z=z)$,
this E_a will apply.