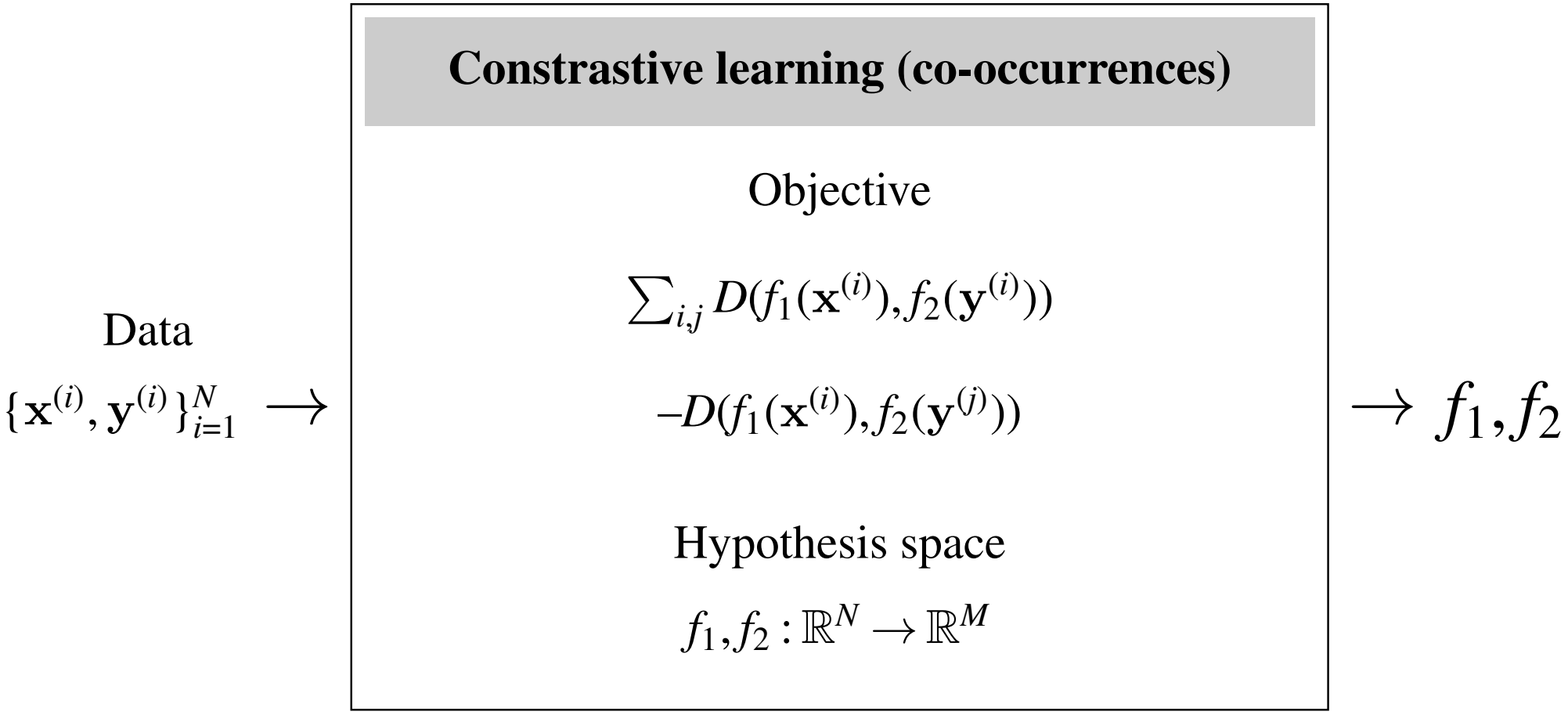
Constrastive learning (co-occurrences)
Data
$\{\mathbf{x}^{(i)}, \mathbf{y}^{(i)}\}_{i=1}^{N}$
Objective
$\sum_{i,j} D(f_1(\mathbf{x}^{(i)}), f_2(\mathbf{y}^{(i)}))$
$-D(f_1(\mathbf{x}^{(i)}), f_2(\mathbf{y}^{(j)}))$
Hypothesis space
$f_1, f_2 : \mathbb{R}^N \to \mathbb{R}^M$
$\to f_1, f_2$