

$$y_j = \begin{cases} 1 & \text{if index } j \text{ is selected} \\ 0 & \text{otherwise} \end{cases}, j=1, \dots, n$$

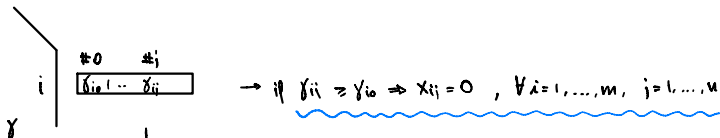
↳ index 0 is selected by default

$$x_{ij} = \begin{cases} 1 & \text{if query } i \text{ uses index } j \\ 0 & \text{otherwise} \end{cases}, i=1, \dots, m, j=0, \dots, n$$

$$\left\{ \begin{array}{l} \min \underbrace{\sum_{i=1}^m c_i y_i}_{\text{fixed cost}} + \underbrace{\sum_{i=1}^m \sum_{j=0}^n d_{ij} x_{ij}}_{\text{query cost}} \\ \sum_{i=1}^m d_i y_i \leq D \\ 0 \leq y_i \leq 1 \text{ integer} \\ \sum_{j=0}^n x_{ij} = 1 \rightarrow \text{w. of selected entries in row } i \\ 0 \leq x_{ij} \leq 1 \text{ integer} \\ \sum_{i=1}^m x_{ij} \leq m y_j \rightarrow \text{Bad} \end{array} \right.$$

This LB is not tight : we must introduce new constraints so the gap gets smaller

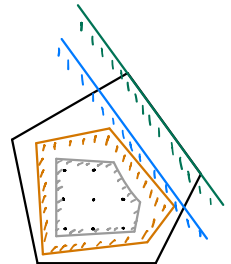
1) Data preprocessing



↳ input data → We can force the variables to be 0



reducing the number of variables ⇒ lower m



Or even better :

Cut C_1 :

$$\underline{x_{ij} \leq y_i} \quad i=1, \dots, m, j=1, \dots, n$$

Cut C_2 :

$$\underline{y_i + y_j \leq 1, \forall ij : d_i + d_j > D, i < j}$$