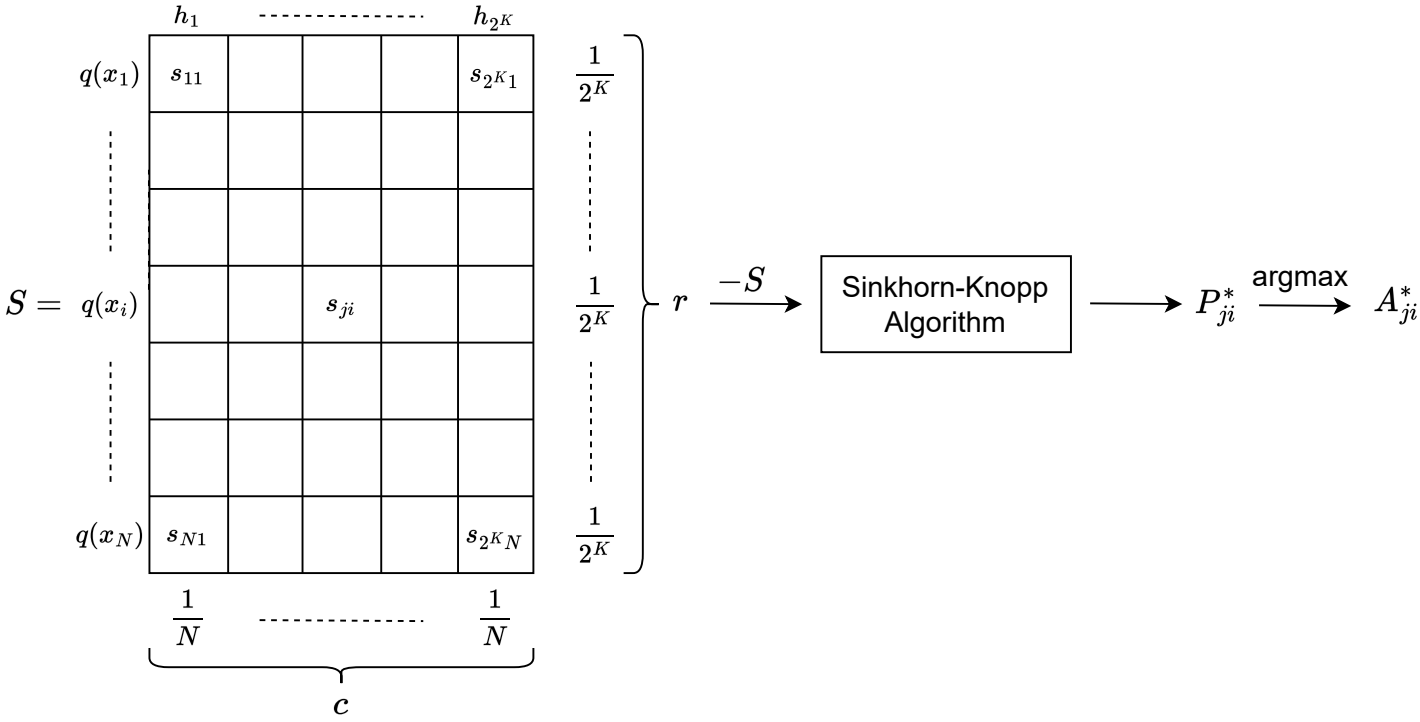

S: cosine similarity between *q(x)* and h_k. *-S* represents transportation cost.

N: total *q(x)* samples in a mini-batch.

K: bits to encode each cluster centroid h_k. 2^K are total clusters.

r, c: marginal distributions as constraints for uniform clustering.

$\boldsymbol{P^*_{ji}}$: the probability that $q(x_i)$ belongs to the j^{th} cluster. Represents optimal transportation plan

$\boldsymbol{A^*_i}$: assignment index. One-hot vector.