What You Should know

- What are the deficiencies of PLSA that have motivated LDA?

- What is the main difference of LDA and PLSA? How many parameters are there in LDA? Can you write down the likelihood function of LDA?

- How to describe PLSA and LDA using the plate notation

- Why do we have to use posterior inference to compute the word distributions and topic coverage in each document in LDA, but we can use Maximum Likelihood estimate to obtain them in PLSA?

$$p(\{\theta_j\}, \{\pi_{d,j}\} \mid C, \hat{\bar{\alpha}}, \hat{\bar{\beta}}) = ?$$