

1. Как работает Snapshot Ensembles? В чем заключается его отличие от Fast Geometric Ensembling?
2. Что такое широкий и узкий минимум функции потерь? Что из этого лучше с точки зрения обобщающей способности? Как работает Stochastic Weight Averaging?
3. Что такое Pruning? Опишите какой-нибудь метод Pruning-a. Чем отличается One-Shot Pruning от Iterative Pruning?
4. Опишите Lottery Ticket Hypothesis. Как найти Lottery Ticket? Как изменится его качество, если сделать в нем некоторую случайную инициализацию и обучить заново?