ANN中，深度和單層但節點數多，哪個好

1024 vs 2\*\*10

overfit ->因為變數太多

eta: learning rate 大小看經驗

cross entropy 越小越準確

soft max: normalization 成機率