



deeplearning.ai

# Hyperparameter tuning

---

## Tuning process

# Hyperparameters

→  $\alpha$

$\beta$  0.9

$\beta_1, \beta_2, \epsilon$   
0.9 0.999  $10^{-8}$

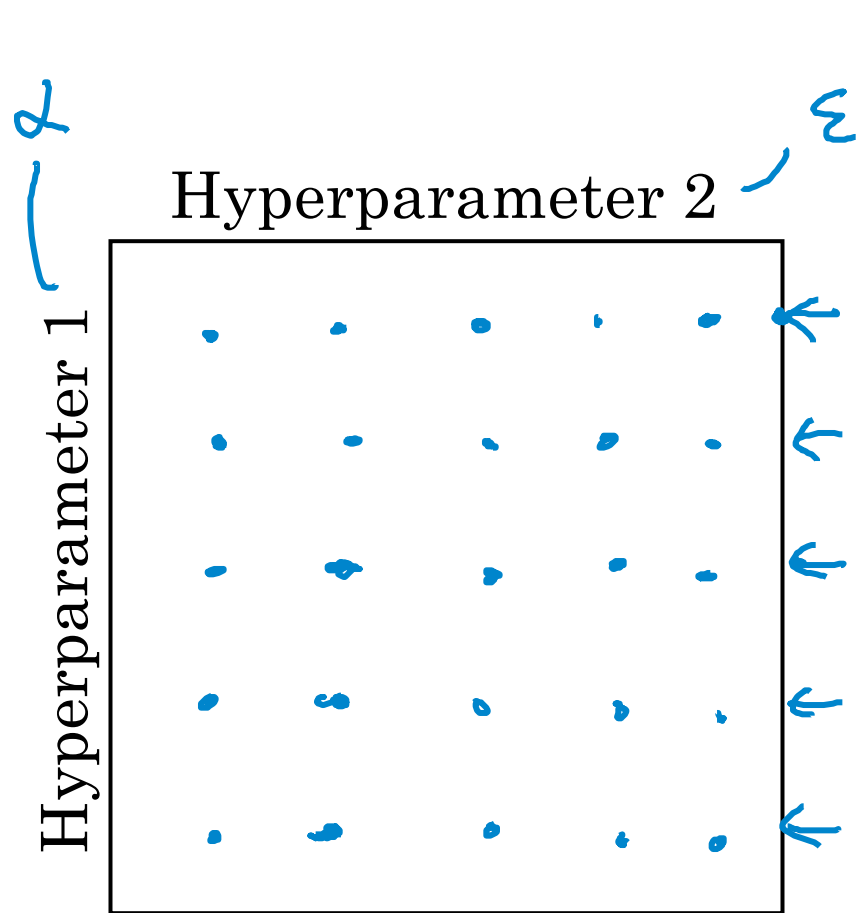
# layers

# hidden units

learning rate decay

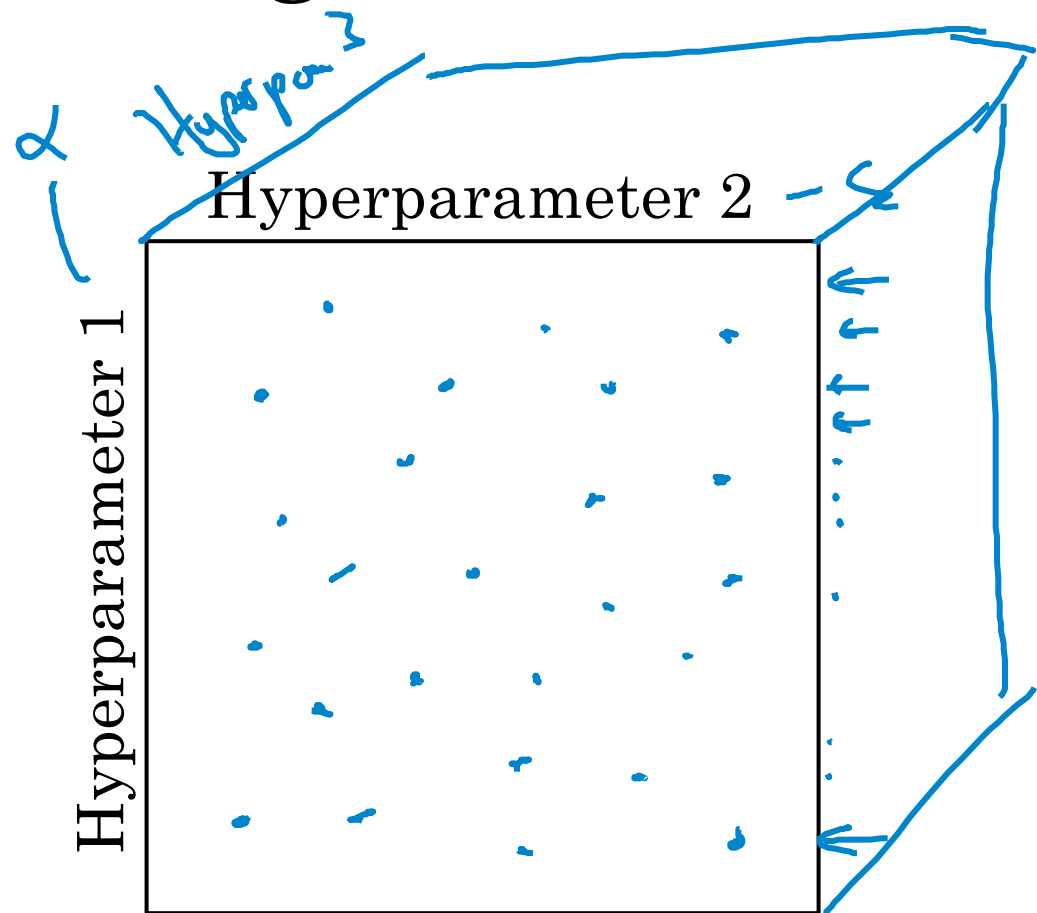
mini-batch size

# Try random values: Don't use a grid



grid

앞에서 본 것처럼 어떤 하이퍼파라미터는 다른것보다 더 중요한데 그것을 미리 알기는 힘들다. alpha와 epsilon이라고 치면 epsilon은 별 영향을 못미치기 때문에 25개의 모델을 학습하더라도 결국 5개의 learning rate에 대한 시도만 테스트하는 셈이다.



random

반면 랜덤을 사용하면 25개의 서로 다른 alpha 값을 테스트해보는 것이기 때문에 (more richly exploring set of possible values for the most important hyperparameters) 더 좋은 값을 찾을 확률이 높아진다.

# Coarse to fine

another common practice는 coarse to fine sampling scheme을 사용하는 것이다.

