using an appropriate scale to pick hyperparameters

- sampling uniformly random
- appropriate scale for hyperparameters
  - `r=-4*np.random.rand()` $\leftarrow$ $r \in [-4, 0]$
  - `alpha = np.power(10, r)`
- sample from 10^a to 10^b randomly
  - $\alpha = 10^r,\ r \in [a, b]$

Hyperparameters for exponentially weighted averages

$\beta = 0.9 \cdots 0.999$

$1 - \beta = 0.1 \cdots\ 0.001$

$r \in [-3, -1],\ 1 - \beta = 10^r,\ \beta = 1 - 10^r$

$\beta\ is\ more\ sensitive\ when\ \beta\ is\ close\ to\ 1$

---

strategy of choosing hyperparameters