GENERALIZATION

Overfitting the model on training data
produces models that generalize poorly,
i.e. do a worse job on test/real data
than a "worse" model (from POV of
training data w.r.t. loss fn) would have.

How to handle?
    — Want a model as simple as possible
    — Segment a test set and evaluate

Test set assumptions:
1. Draw independently and identically
    (i.i.d.) and randomly from dist.

(user behavior is seasonal) ⟹ 2. Dist. is stationary, doesn't change over time

(e.g. fashion changes) ⟹ 3. Always from same distribution.

The less complex a model is, the more likely
it is that a good empirical result is not just
due to the peculiarities of the sample.