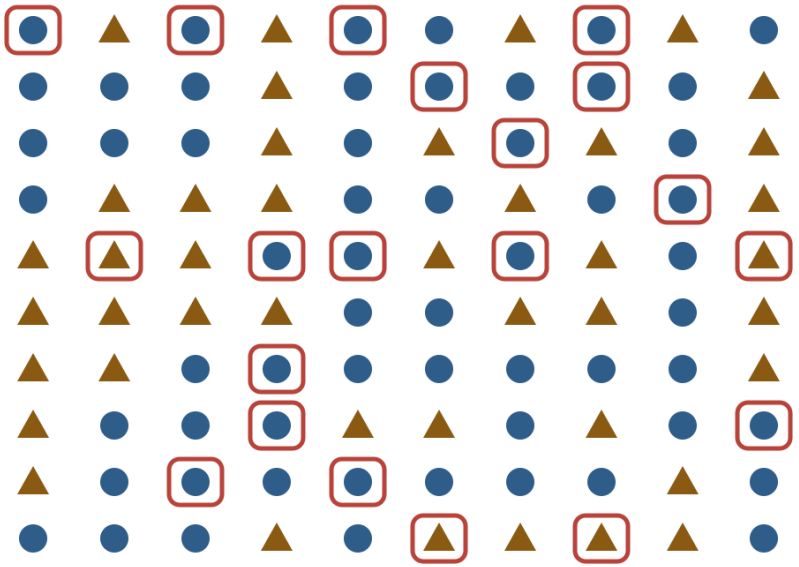
学習用データと確かめ用データ（分け方）

集めたデータを、学習に使うものと確かめに使うものに、どう分ければよいのだろう

集めたデータ（100件）を分ける



●・▲：データの種類（例：ねこ・いぬ）
赤い枠の20件を、テスト用にくじのように選んだ（架空）

分け方の3つのきまり

1 テストデータを学習に使わない

答えを見てからテストを受けるのと同じ。
本当の力が分からなくなる

2 かたよらないように分ける

くじのように選び、種類の割合を
そろえる（ねこだけ、などにしない）

3 本番に近いデータで確かめる

使う場面と同じような写真・条件の
データでテストする

分けた後の使い方

訓練データ（80件）

学習に使う

テストデータ（20件）

最後に一度だけ確かめる

もっと詳しく

訓練・検証・テストの3つに分け、
検証データでモデルの設定を選び、
テストデータは最後に使うこともある。

※ 模式図。データの数と分け方の割合は例。