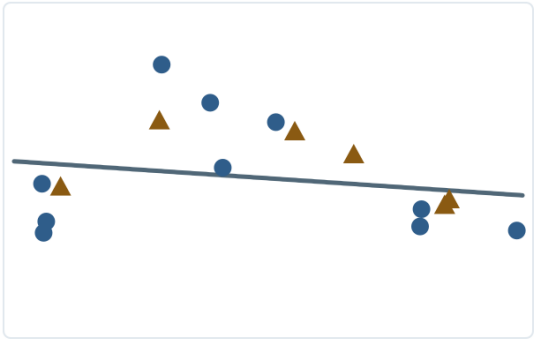
過学習（練習問題だけ覚えて本番で失敗する）

学習に合わせすぎたAIは、なぜ新しいデータでまちがえるのだろう

集めたデータを分ける：　● 訓練データ（学習に使う）　▲ テストデータ（学習に使わず、確かめに使う）

適合不足（単純すぎ）

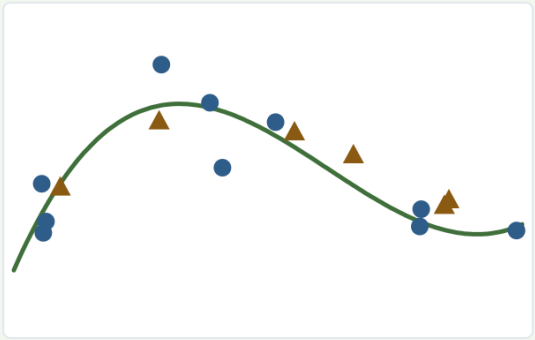


訓練データでの正しさ：低い

テストデータでの正しさ：低い

線が単純すぎて、どちらにも合わない

ちょうどよい

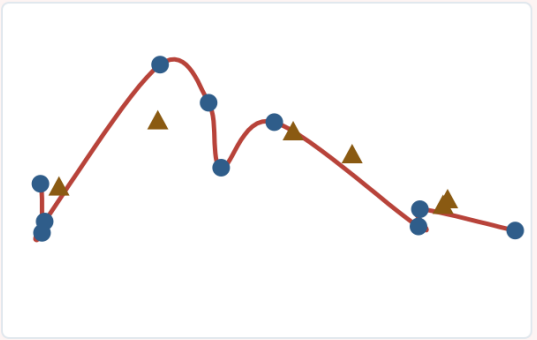


訓練データでの正しさ：高い

テストデータでの正しさ：高い

全体の傾向をとらえている

過学習（過剰適合）



訓練データでの正しさ：とても高い

テストデータでの正しさ：低い

点に合わせすぎ、新しい点を外す

たとえると

練習問題の答えを丸暗記すると、練習では満点。
でも、本番で少し問題が変わると解けない。
→ 新しいデータでも正しく答えられるかが大事

防ぐ工夫

- テストデータで確かめる（訓練データと分ける）
- 多様なデータをふやす
- モデルを必要以上に複雑にしない

※ 模式図。点は架空のデータ（乱数で作った例）。線は1次・3次の式の当てはめと、訓練データのすべての点を通る曲線。