



## 学習不足と過学習

次数を  $d$  とする多項式関数による多項式回帰を考える

ref: なっとく！機械学習 p84~88

$d = 1$  の場合、これは線形回帰であり、関数のグラフは直線になる

ref: 線形代数の半歩先 p133~134


$d > 1$  の場合、関数のグラフは、最大で  $d - 1$  回カーブする曲線になる

### 学習不足（次数が小さすぎる場合）

たとえば、データ点がまったく直線状に並んでいるとはいえないデータセットに対して、うまく適合する直線を見つけることはできない

複雑なデータに  $d = 1$  の線形回帰を適用しても、その結果得られた直線では、未知のデータを予測することはできない

これが学習不足（**underfitting**）の例であり、データセットが複雑であるにもかかわらず、単純なモデルしかないという状態に陥っている

 **学習不足** 訓練データのパターンをうまく認識するにはモデルの複雑さが十分ではなく、未知のデータにうまく汎化（対応）できないこと

学習不足が起こるのは、「単純すぎる」モデルを訓練しようとしたときである


### 過学習（次数が大きすぎる場合）

単純なデータセットに対して、次数  $d$  の大きい多項式回帰を適用した場合にも問題が生じる

グラフのカーブが多すぎると、どのデータ点にもうまく接しているように見えるが、データが存在しないところで余分なカーブが発生する

このような曲線はデータの本質を捉えておらず、未知のデータに対して的外れな予測値を返す可能性が高い

これが過学習（overfitting）の例であり、モデルが過度に柔軟（複雑）だと、データがないところでの振る舞いが不自然になってしまう

 **過学習** モデルは訓練データのパターンを認識しているが、モデルが複雑すぎるために、未知のデータにうまく汎化できないこと

過学習が起こるのは、「複雑すぎる」モデルを訓練しようとしたときである