パターン認識と機械学習
1.1 例:多項式曲線フィッティング

茨城大学工学部情報工学科
佐々木稔

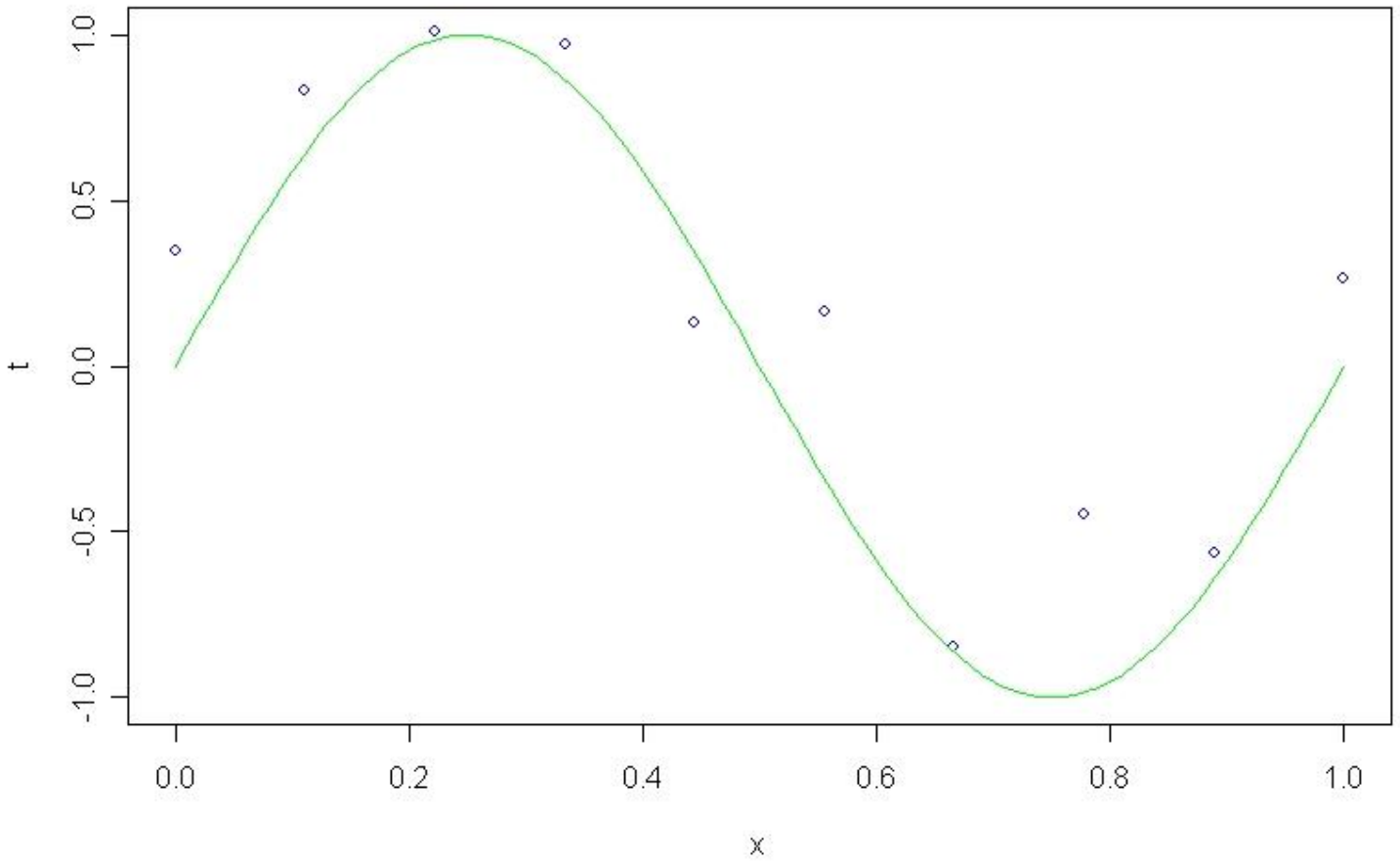
目的

- 実数値の入力変数 x から、
  実数値の目標変数 t を予測したい
- x から t を発生するモデルを予測
  - 訓練集合を使ってモデルへ汎化
  - ノイズの影響で、予測値に不確実性がある
    - 確率論： 不確実性の厳密かつ定量的な表現
    - 決定理論： 適切な基準下で最適な予測をする

データ

- 訓練集合からデータ生成モデルを予測
  - N 個の観測値 $x = (x_1, x_2, \cdots, x_N)^{\mathrm{T}}$
  - 対応する観測値 $t = (t_1, t_2, \cdots, t_N)^{\mathrm{T}}$
- 本節では、人工データを使用
  - $N = 10$
  - 観測値 t は $\sin(2\pi x) + N(0, 0.3)$

訓練集合のグラフ



曲線フィッティング

- 多項式関数（線形モデル）へのあてはめ

$$y(x, \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + \cdots + w_M x^M = \sum_{i=0}^{M} w_i x^i$$

- $\mathbf{w}$：多項式の係数
- M：多項式の次数

係数 w の値の計算

- 誤差関数の最小化
  - w を任意に固定した時の $y(x, \mathbf{w})$ と
    訓練集合のデータ点との誤差を最小にする
- 誤差関数：最小二乗誤差

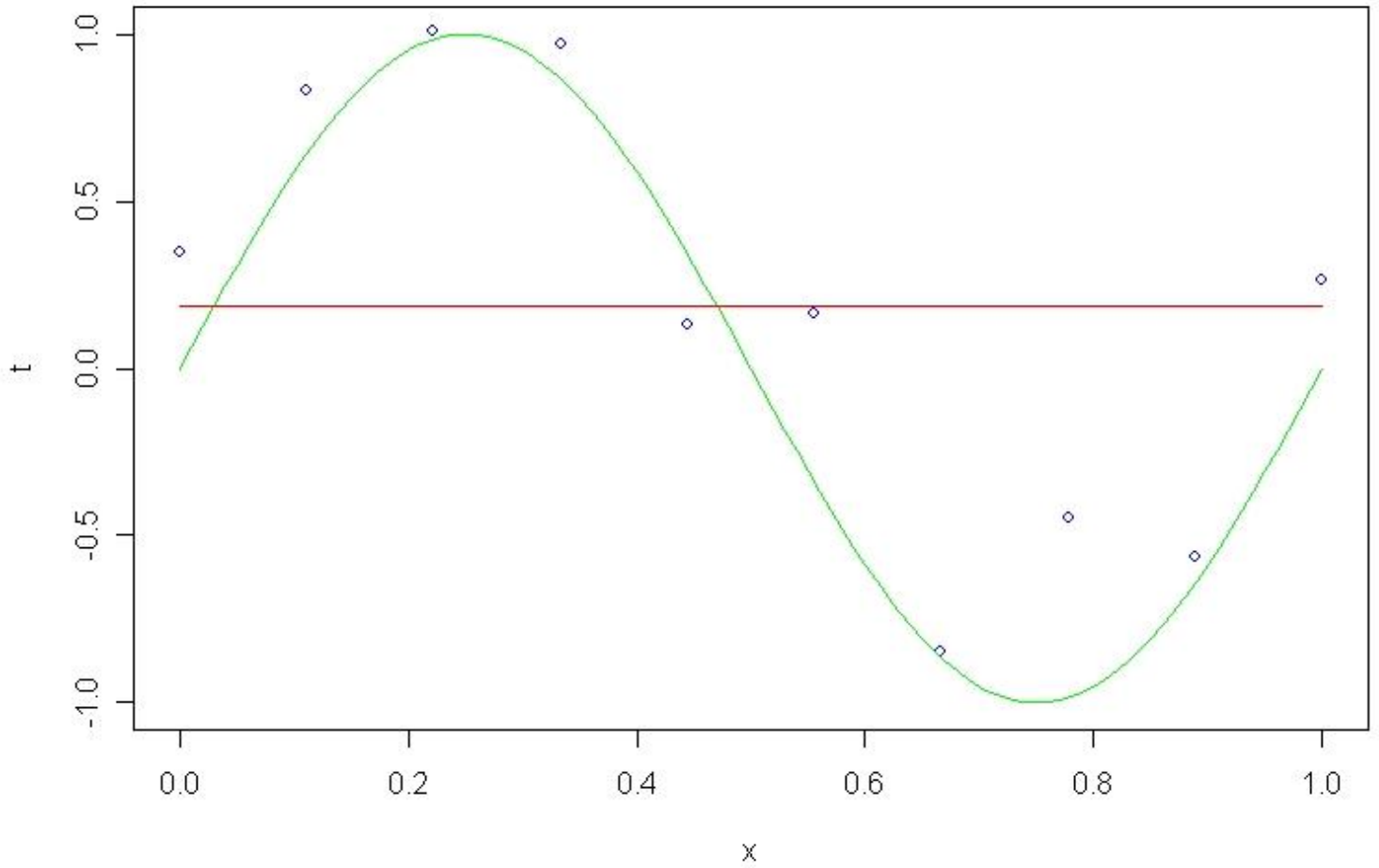
$$E(\mathbf{w}) = \frac{1}{2}\sum_{n=0}^{N}\{y(x_n, \mathbf{w}) - t_n\}^2$$

- $E(\mathbf{w})$ を最小にする w を選ぶ

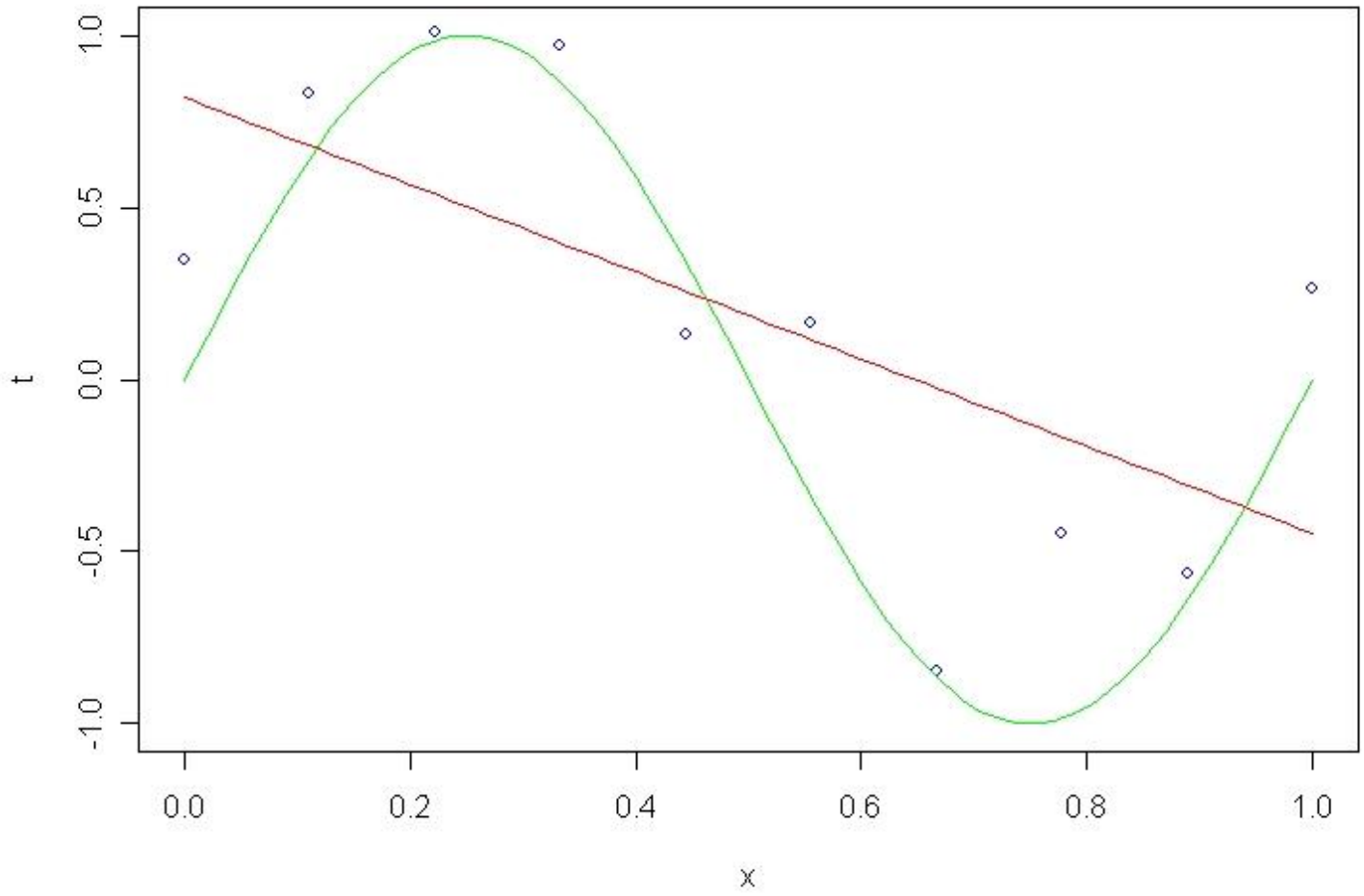
モデル選択

- 多項式の次数 M の選択
- M=0, 1 の場合
  - 当てはまらないので不適当
- M=3, 6 の場合
  - 当てはまっているように見える
- M=9 の場合
  - うまく当てはまっているが振動の大きい曲線
  - 関数 $\sin(2\pi x)$ とは異なる → 過学習

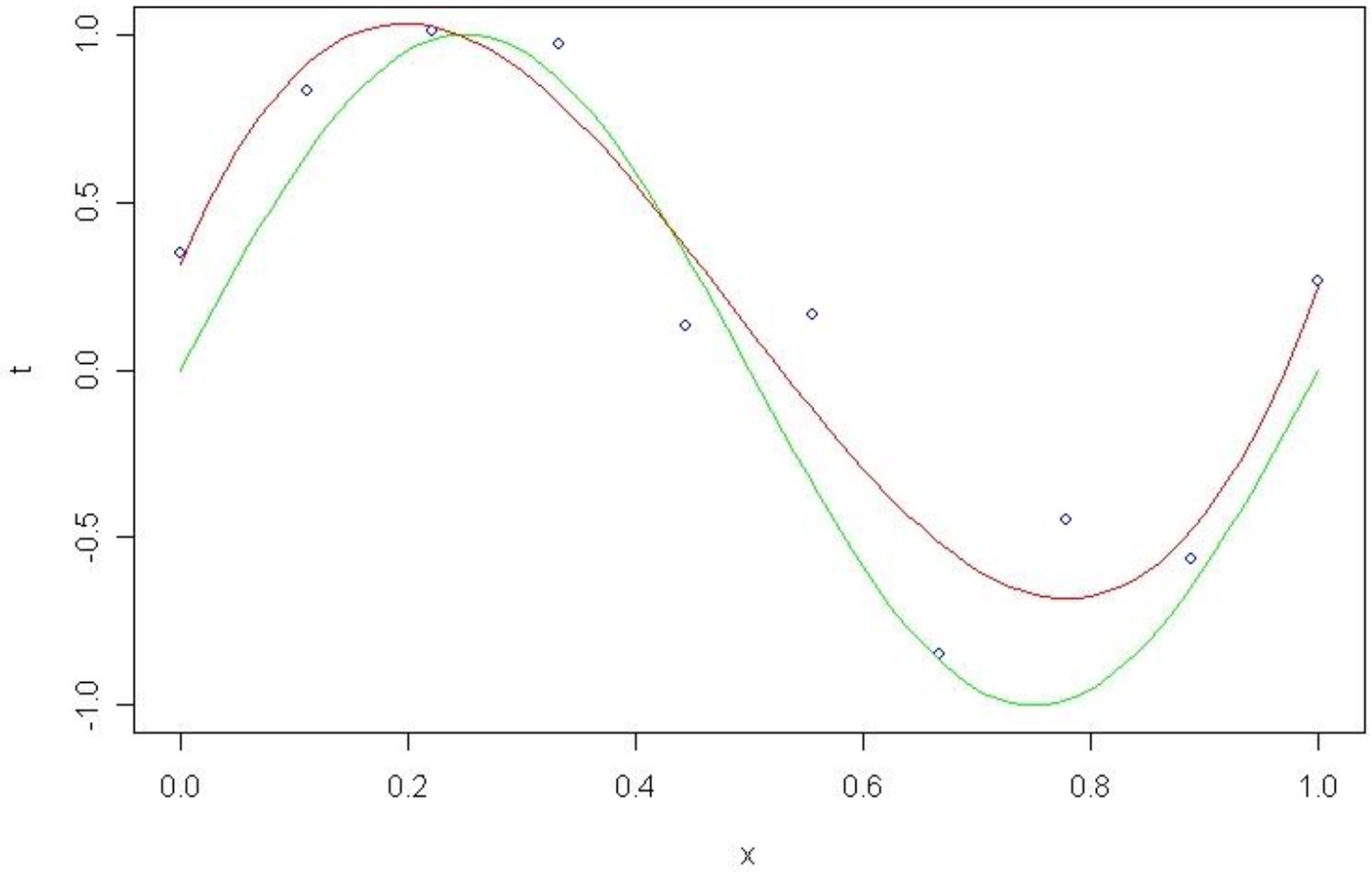
M=0 における予測値



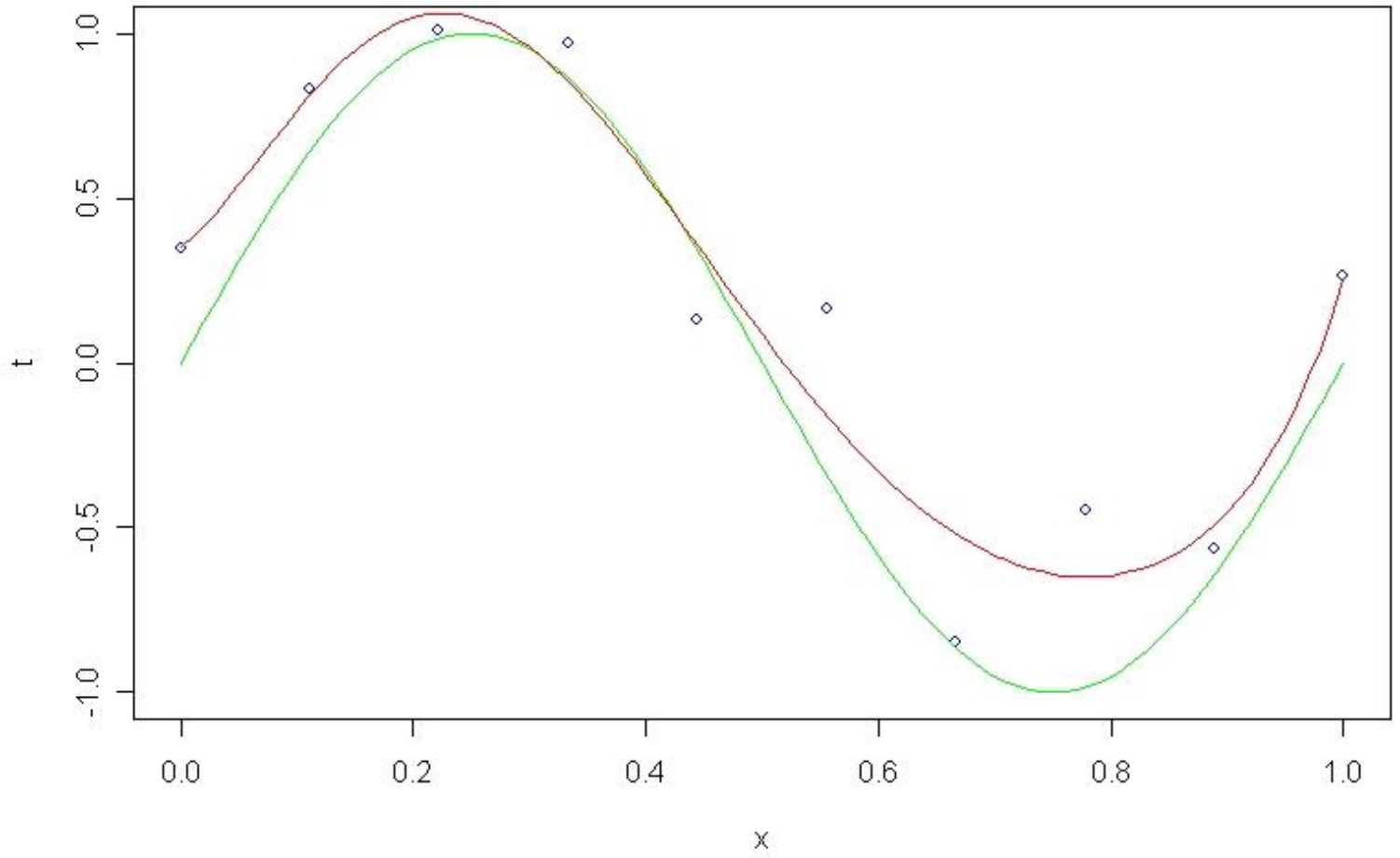
$M=1$ における予測値



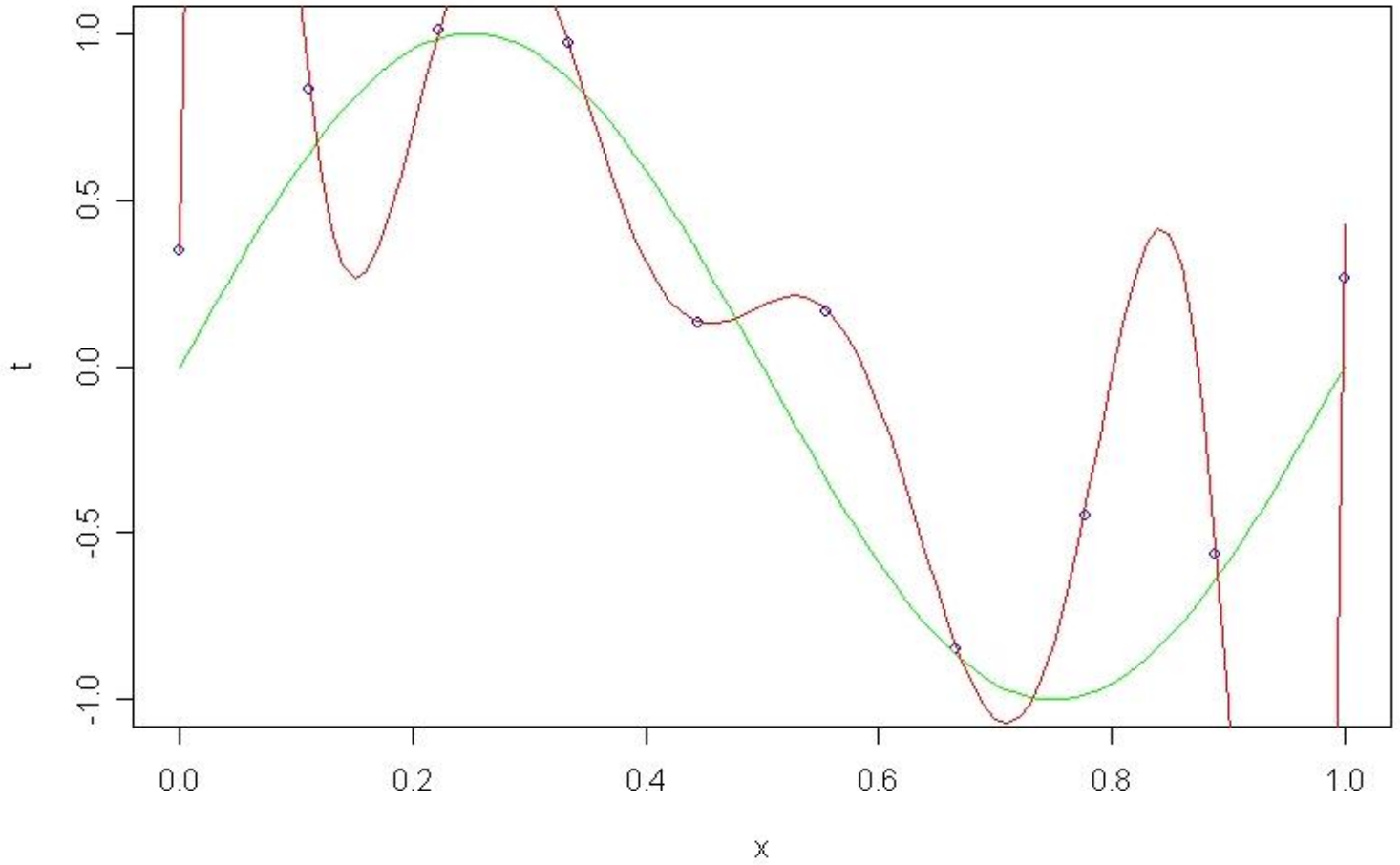
M=3 における予測値



M=6 における予測値



M=9 における予測値

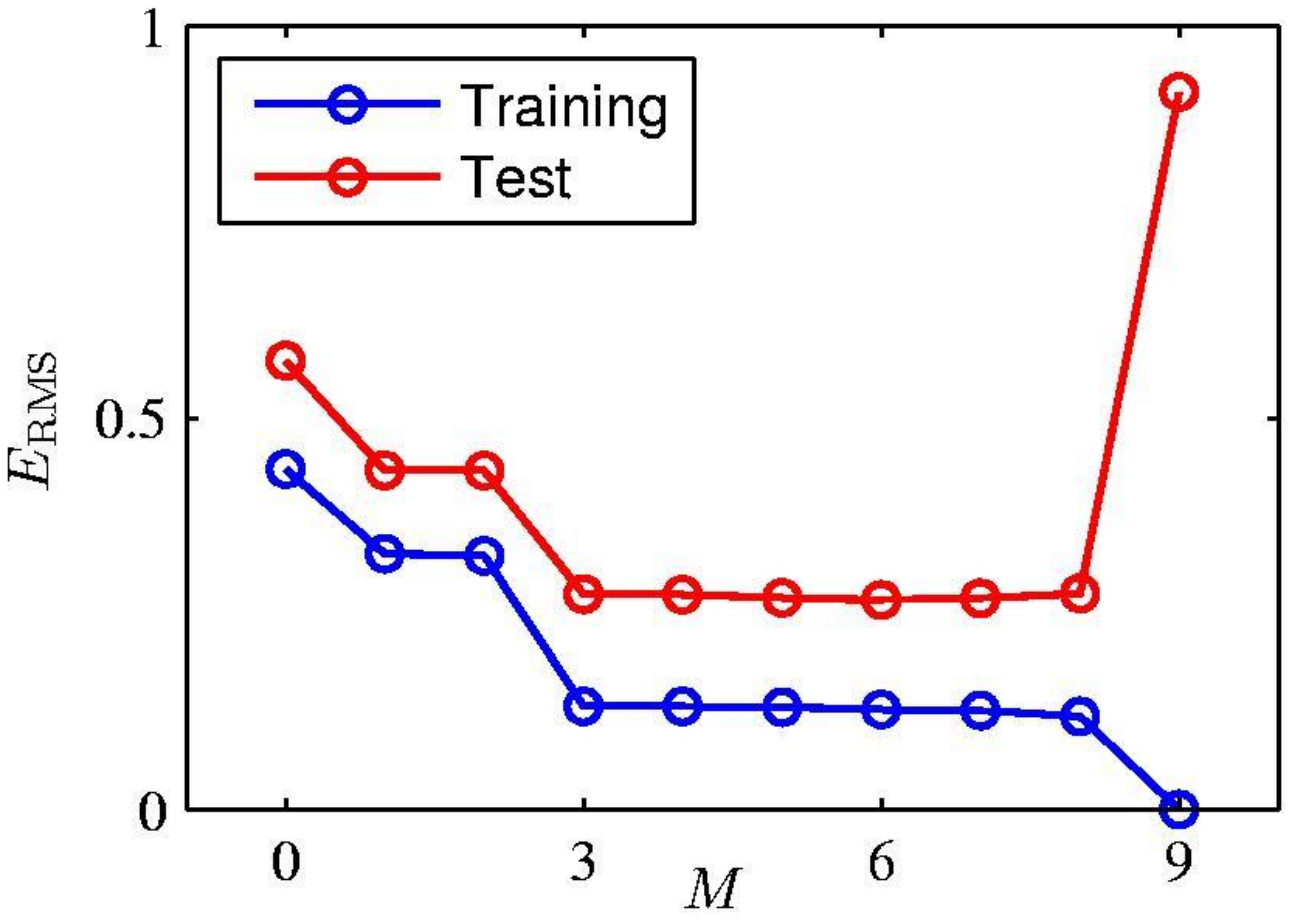


テストデータからのモデル精度

- M の違いによる汎化性能を定量的に評価
- 訓練集合で求めた $\mathbf{w}^*$ の評価
  - 独立したテストデータ 100 個用意
  - $\mathbf{w}^*$ による予測値との誤差 $E(\mathbf{w}^*)$ を計算
- 評価尺度: 平均二乗平方根誤差

$$E_{\mathrm{RMS}} = \sqrt{\frac{2E(\mathbf{w}^*)}{N}}$$

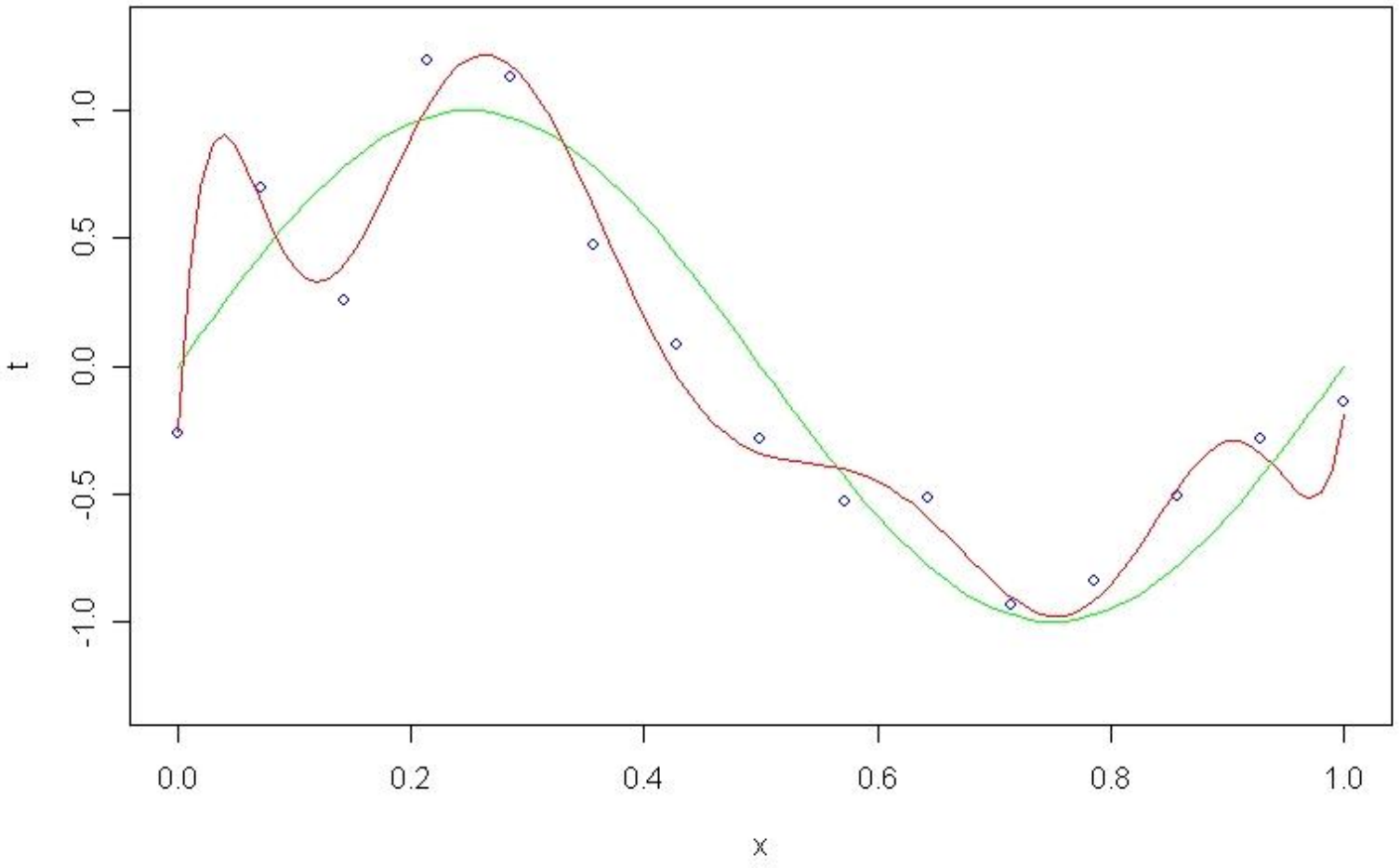
次数 M に対するE_{RMS}の値



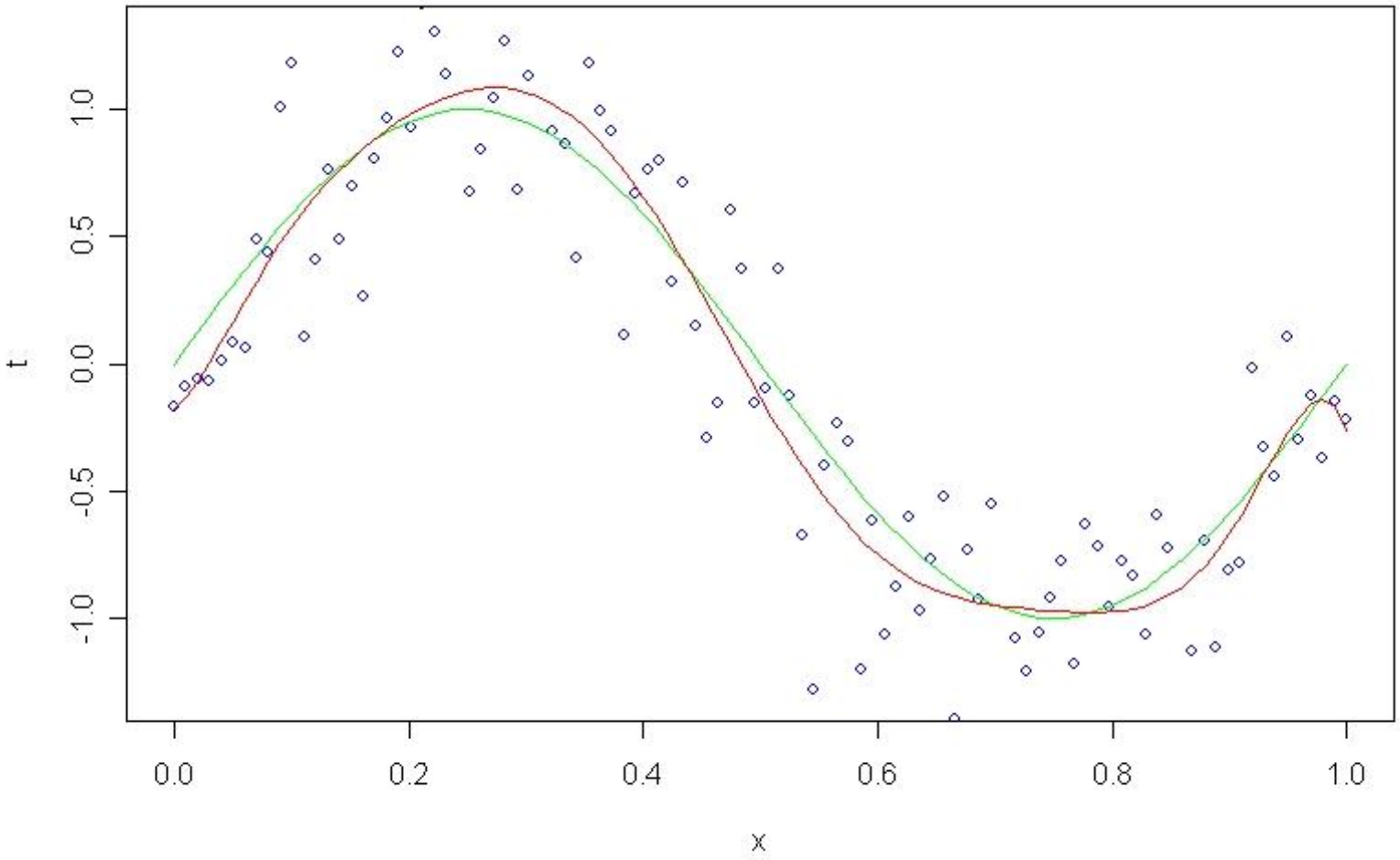
訓練集合のサイズ

- 次数は固定し、データの数を変更
  - 次数 $M = 9$
  - データ数 15 と 100

N=15におけるグラフ



N=100におけるグラフ



データ数の変化と精度

- 曲線の発振が小さくなる
- データ数の増加で、予測値が近くなる
- データ数の目安
  - パラメータ数の5倍は必要
- モデルの複雑さを測る尺度
  - パラメータ数だけではない
  - データ数も影響

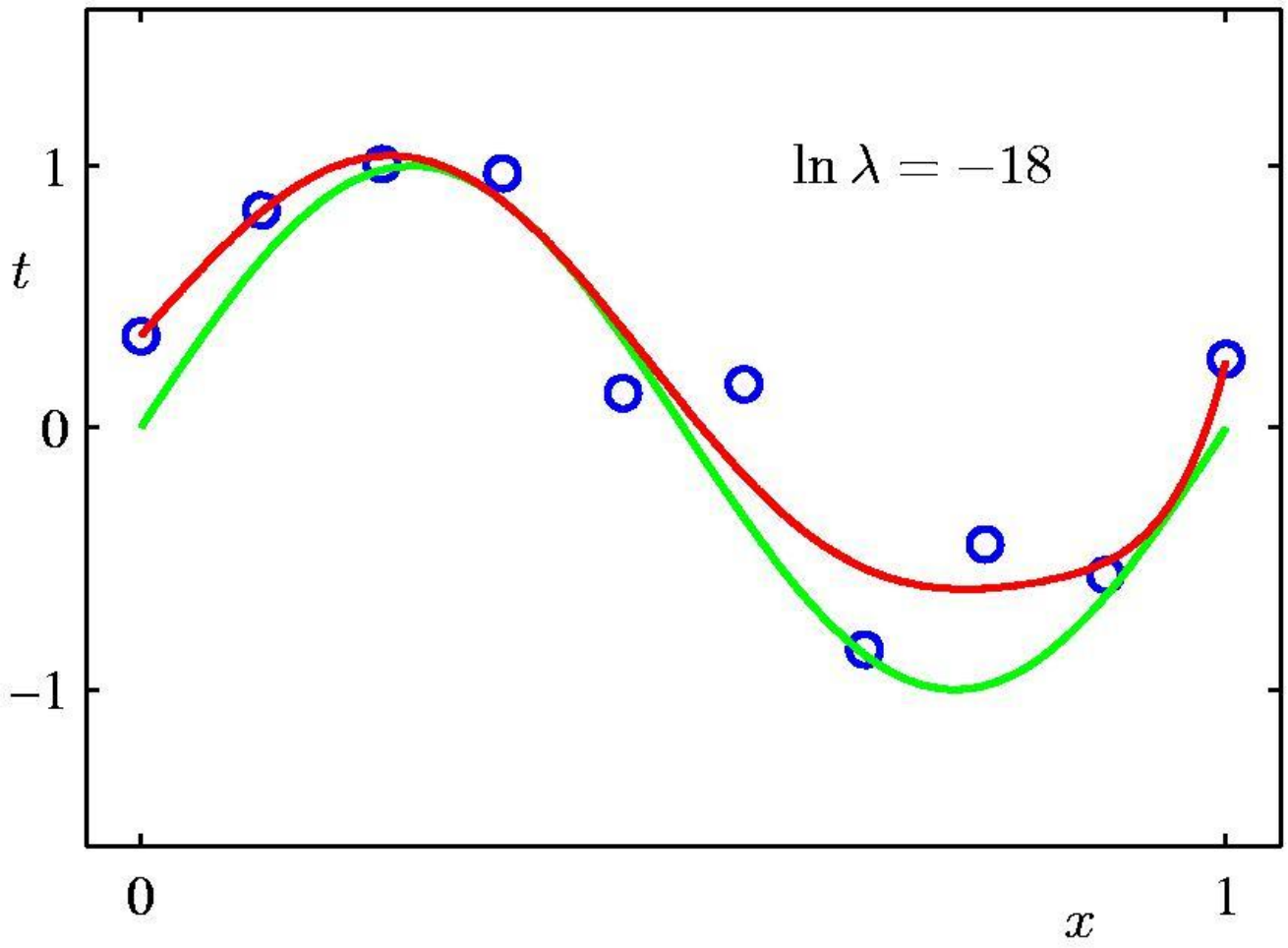
正則化

- 限られたデータ数で良いモデルを求めたい
  - 過学習を抑制する
- 正則化
  - 誤差関数にペナルティ項を付加

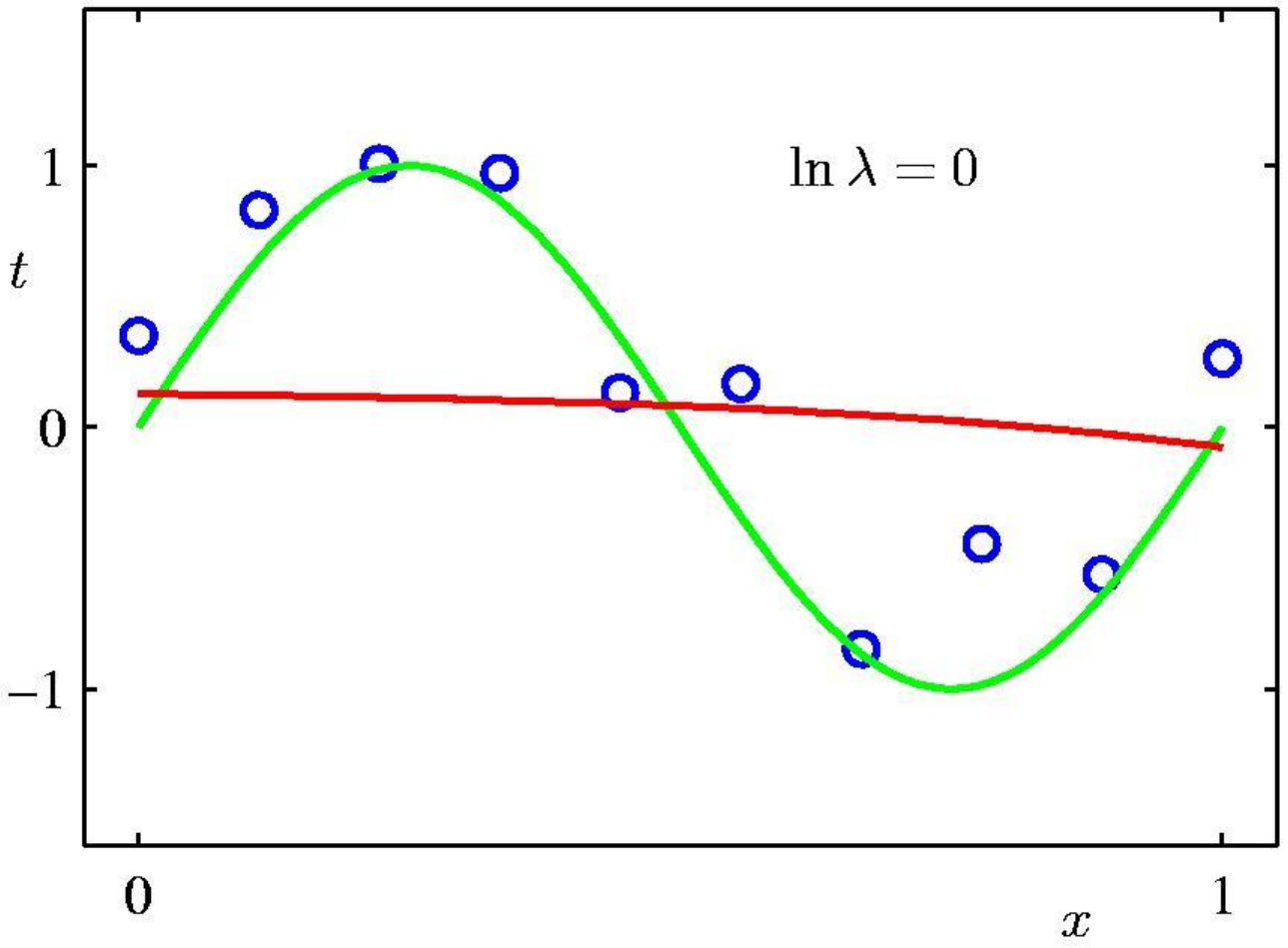
$$\tilde{E}(\mathbf{w}) = \frac{1}{2}\sum_{n=0}^{N}\{y(x_n, \mathbf{w}) - t_n\}^2 + \frac{\lambda}{2}\|\mathbf{w}\|^2$$

$$\|\mathbf{w}\|^2 = w_0^2 + w_1^2 + \cdots + w_M^2$$

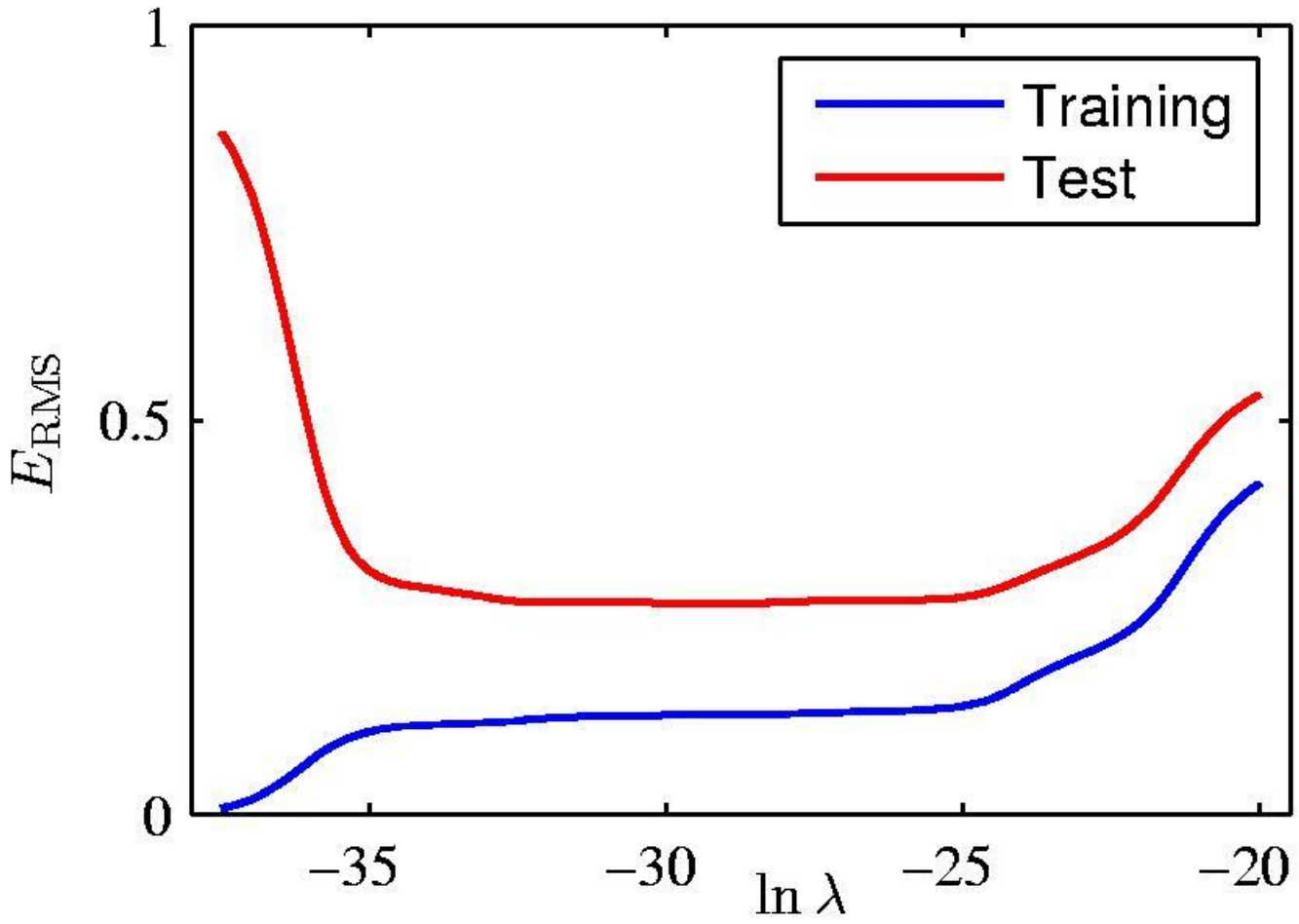
Ln λ=-18 におけるグラフ



Ln λ = 0 におけるグラフ



正則化項による誤差の影響



モデルの複雑さ

- 誤差関数の最小化
  - モデルの複雑さを決める方法が必要
- 方法
  - データを訓練集合と確認用集合に分ける
    - 訓練集合の一部が無駄になることがある
    - より洗練されたアプローチを見つける必要

まとめ

- 最小二乗法による多項式関数への回帰
  - 次数を大きくすると振動が大きくなる
  - 別のテストデータでパラメータの評価が可能
  - データ数の増加で振動が少なくなる
  - 正則化項をつけると振動が少なくなる
  - モデルの複雑さを決めるのは難しい問題