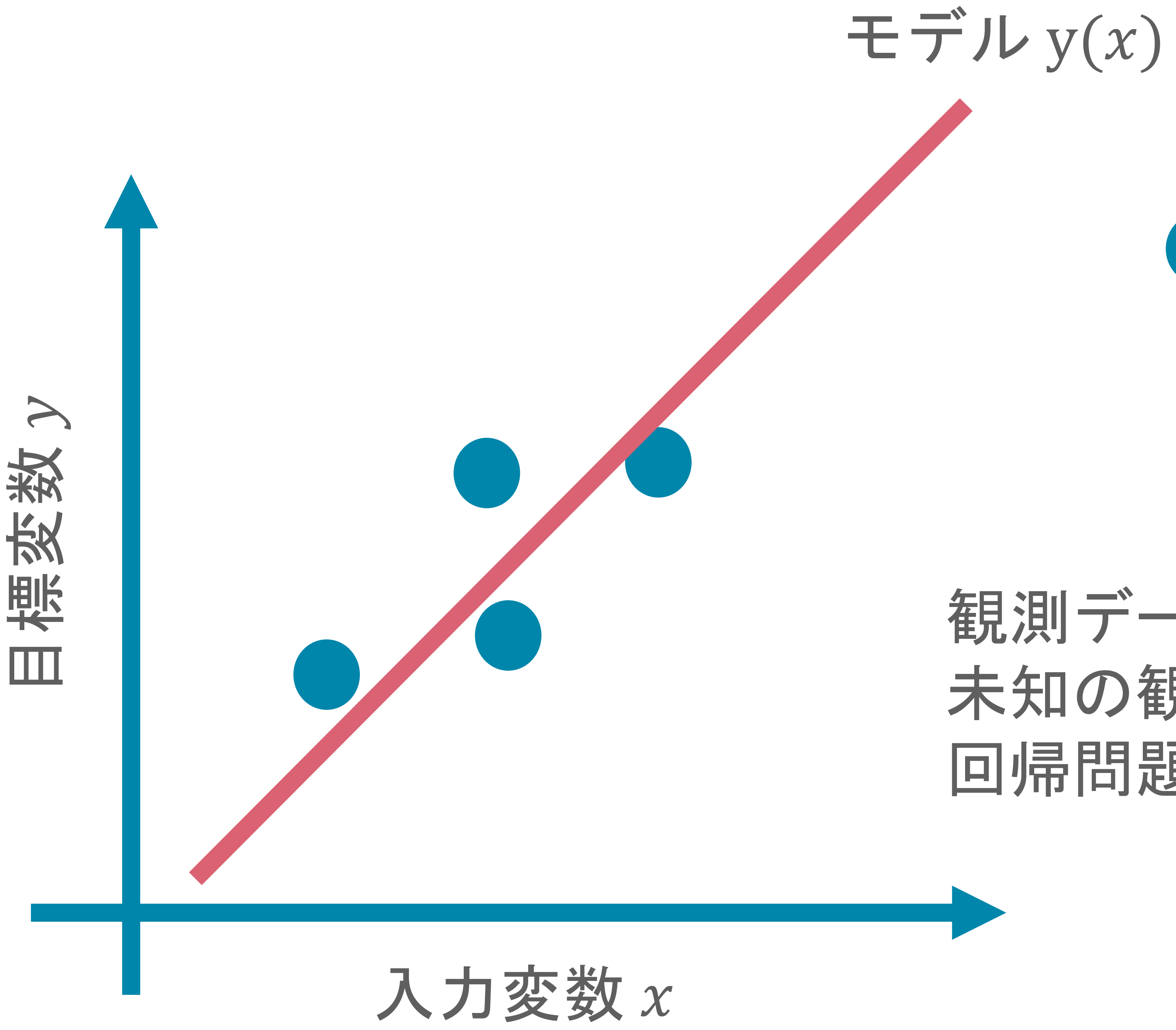
ベイズ推定に基づく線形回帰

知能システム制御研究室
学籍番号2011115 佐伯 雄飛

目次

背景

回帰問題とは



● 観測データ $\mathrm{D} = \{(x_i, y_i); i = 1,2, \ldots, n\}$

観測データDから，**回帰関数モデル$\boldsymbol{y}(\boldsymbol{x})$**を推定し，
未知の観測データxに対する**出力$\boldsymbol{y}$**を予測することが，
回帰問題の目的となる．

回帰関数モデル

回帰関数モデルは基底関数の線形結合に基づき，以下のように定義される

$$y(x) = \Phi(x) * w + \varepsilon \quad (1)$$

$\Phi(x)$: xの基底関数

w : 基底関数の係数　　←ここを推定する

ε : 誤差項

推定方法

最小二乗推定

方針：**予測誤差の二乗和$S(w)$を最小化する$\hat{w}$を推定値とする**

$$S(w) = \varepsilon^T \varepsilon = (y - \Phi w)^T (y - \Phi w) \quad (2)$$

$S(w)$をwで偏微分して$\hat{w}$を求める．

$$\frac{dS(w)}{dw} = -\Phi^T y + \Phi^T \Phi w$$

$\frac{dS(w)}{dw} = 0$ のとき

$$\hat{w} = (\Phi^T \Phi)^{-1} \Phi^T y$$

従って，最小二乗推定による予測モデルは，

$$\hat{y} = \Phi \hat{w} = \Phi (\Phi^T \Phi)^{-1} \Phi^T y$$

最尤推定

方針：**尤度$\boldsymbol{P(y|w)}$**を最大化する$\hat{w}$を推定値とする

観測データDは何かしらの分布Hに基づいて生成されると考え，
データが発生する確率(尤度)が最大化する場合を考える.

誤差項に正規分布を仮定した際，
観測値yは，平均Φw，分散行列$\sigma^2 In$のn次元正規分布に従う.
よって，尤度は

$$P(y|w) = \mathcal{N}(\Phi w, \sigma^2) \quad (3)$$

→最大化

最尤推定

$$P(y|w) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2\sigma^2}(y-\Phi w)^T(y-\Phi w)\right\}$$

両辺の対数を取る

$$logP(y|w) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{(y-\Phi w)^T(y-\Phi w)}{2\sigma^2}$$

wで微分する

$$\frac{1}{P(y|w)}\frac{P(y|w)}{dw} = -(\Phi^T y + \Phi^T \Phi w)$$

最尤推定

$\frac{P(y|w)}{dw} = 0$のときを考えると

$$\hat{w} = (\Phi^T \Phi)^{-1} \Phi^T y$$

従って，最尤推定による予測モデルは,

$$\hat{y} = \Phi \hat{w} = \Phi (\Phi^T \Phi)^{-1} \Phi^T y$$

(最小二乗推定と同じ結果になる)

MAP推定

方針：wの事後分布$\boldsymbol{P(w|y)}$を最大化する$\hat{w}$を推定値とする

wを確率変数として扱いベイズ推定により事後分布$P(w|y)$を求める.

ベイズの定理

$$\underbrace{P(w|y)}_{\text{事後分布}} = \frac{\overbrace{P(y|w)}^{\text{尤度}}\underbrace{P(w)}_{\text{事前分布}}}{P(y)}$$

MAP推定

事前分布と尤度分布は以下のものを導入する.

$$P(w) = \mathcal{N}(0, \alpha^{-1} I_n) \quad (4)$$

$$P(y|w) = \mathcal{N}(\Phi w, \beta^{-1} I_n) \quad (5)$$

よって，事後確率は以下のように表せる.

$$P(w|y) = \frac{\frac{1}{(2\pi\beta^{-1})^{\frac{n}{2}}} \exp\{-\frac{1}{2\beta^2}(y\Phi w)^T(y - \Phi w)\} \frac{1}{(2\pi\alpha^{-1})^{\frac{n}{2}}} \exp\{-\frac{1}{2\alpha^2} w^T w\}}{P(y)}$$

$P(w|y)$を最大化する$\hat{w}$は，分子(Zとおく)を最大化する$\hat{w}$と等しい

$$\frac{dZ}{dw} = -\frac{1}{2\beta^{-1}}(-\Phi^T + \Phi^T \Phi w) - \frac{1}{2\alpha^{-1}} w^T w$$

MAP推定

$\frac{dZ}{dw} = 0$のときを考えると，

$$\hat{w} = \left(-\frac{\alpha}{\beta} I_n + \Phi^T \Phi\right)^{-1} \Phi^T y$$

従って，MAP推定による予測モデルは，

$$\hat{y} = \Phi\hat{w} = \Phi\left(-\frac{\alpha}{\beta} I_n + \Phi^T \Phi\right)^{-1} \Phi^T y$$

ベイズ推定

方針：wの**事後分布**$\boldsymbol{P(w|y)}$を確率分布として区間推定する

MAP推定では，$P(w|y)$ の最大化を考えたため $P(y)$ を無視したが，
ベイズ推定では考える必要がある．

$P(y)$を周辺化により求める．

$$P(y) = \int P(y|w)P(w)dw$$

$$P(w|y) = \frac{P(y|w)P(w)}{\int P(y|w)P(w)dw}$$

ベイズ推定

ガウス分布に対するベイズの定理より，

$$P(w|y) = \mathcal{N}(u_N, \Sigma_\mathrm{N}) \ \ (6)$$

参考文献PRML p90 ガウス分布と周辺分布と条件付き分布を用いて計算する

$$u_N \left(\frac{\alpha}{\beta} I_n + \Phi^T \Phi\right)^{-1} \Phi^T y \ (7)$$

$$\Sigma_N = (\alpha I_n + \beta \Phi^T \Phi)^{-1} \ (8)$$

従って，ベイズ推定による予測モデルは,

$$\hat{y} = \Phi \widehat{w} = \Phi \left(-\frac{\alpha}{\beta} I_n + \Phi^T \Phi\right)^{-1} \Phi^T y$$

$$\Sigma_N = (\alpha I_n + \beta \Phi^T \Phi)^{-1}$$

数値実験

概要

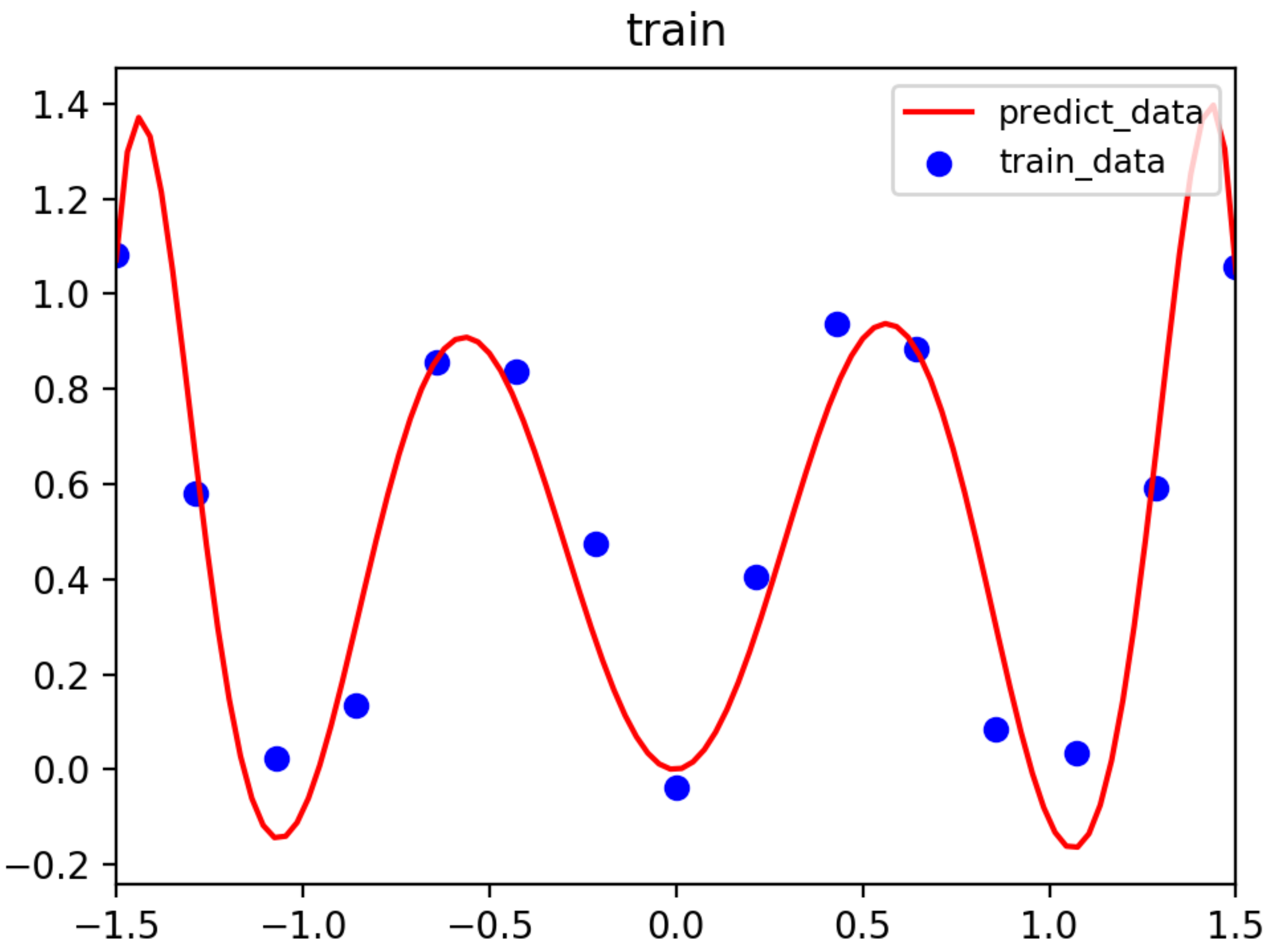
$D = \{(x.train, y.train); i = 1,2, \dots, 15\}$の訓練データに対して，
最小二乗推定，最尤推定，MAP推定，ベイズ推定を用いてモデルを作成
$D = \{(x.test, y.test); i = 1,2, \dots, 100\}$ のテストデータに対して予測分布を確認した．

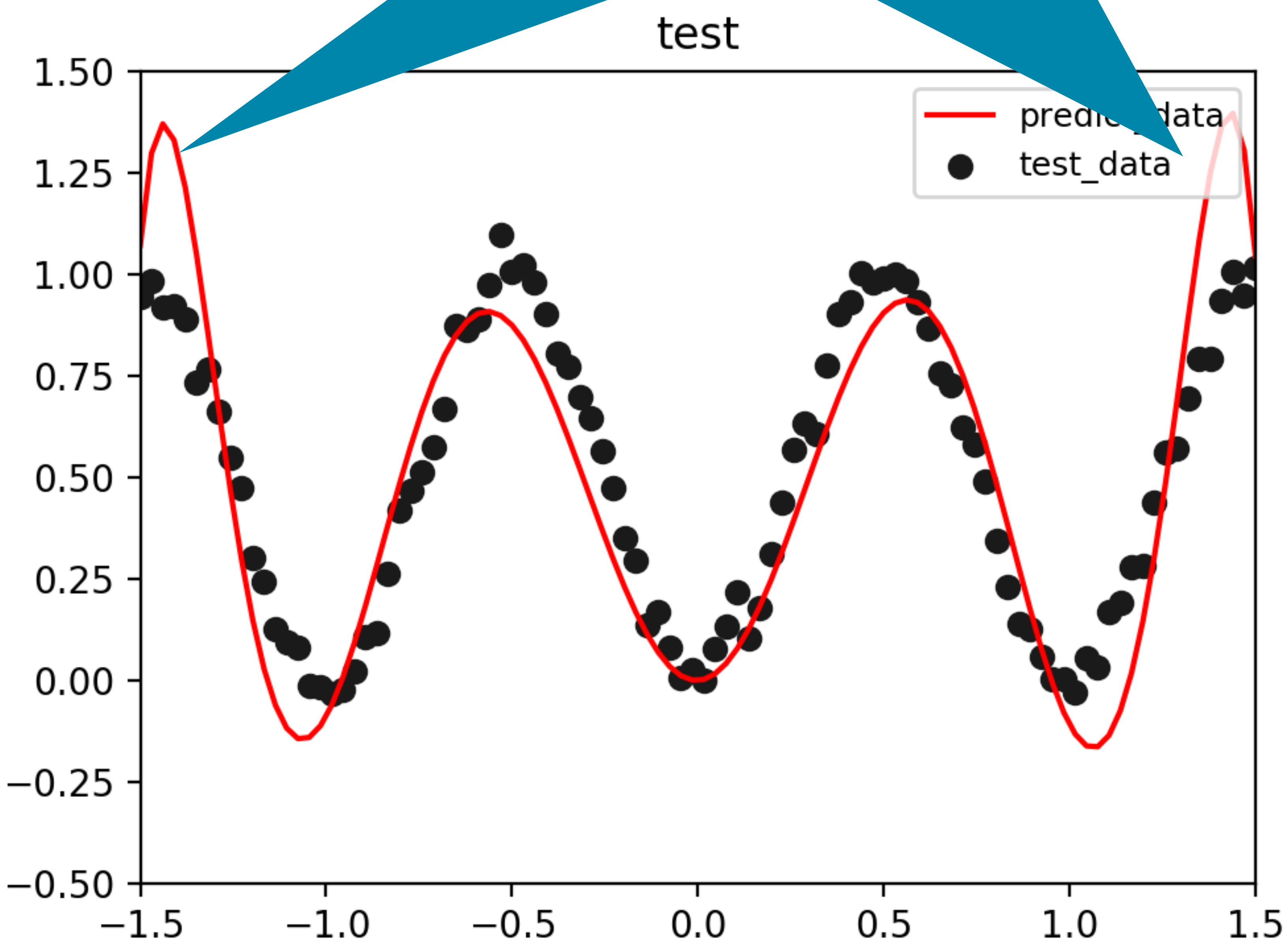
基底関数は以下のものを用いる．

$$f_j(x) = x^j, j = 0,1, \dots, 9 \quad (10)$$

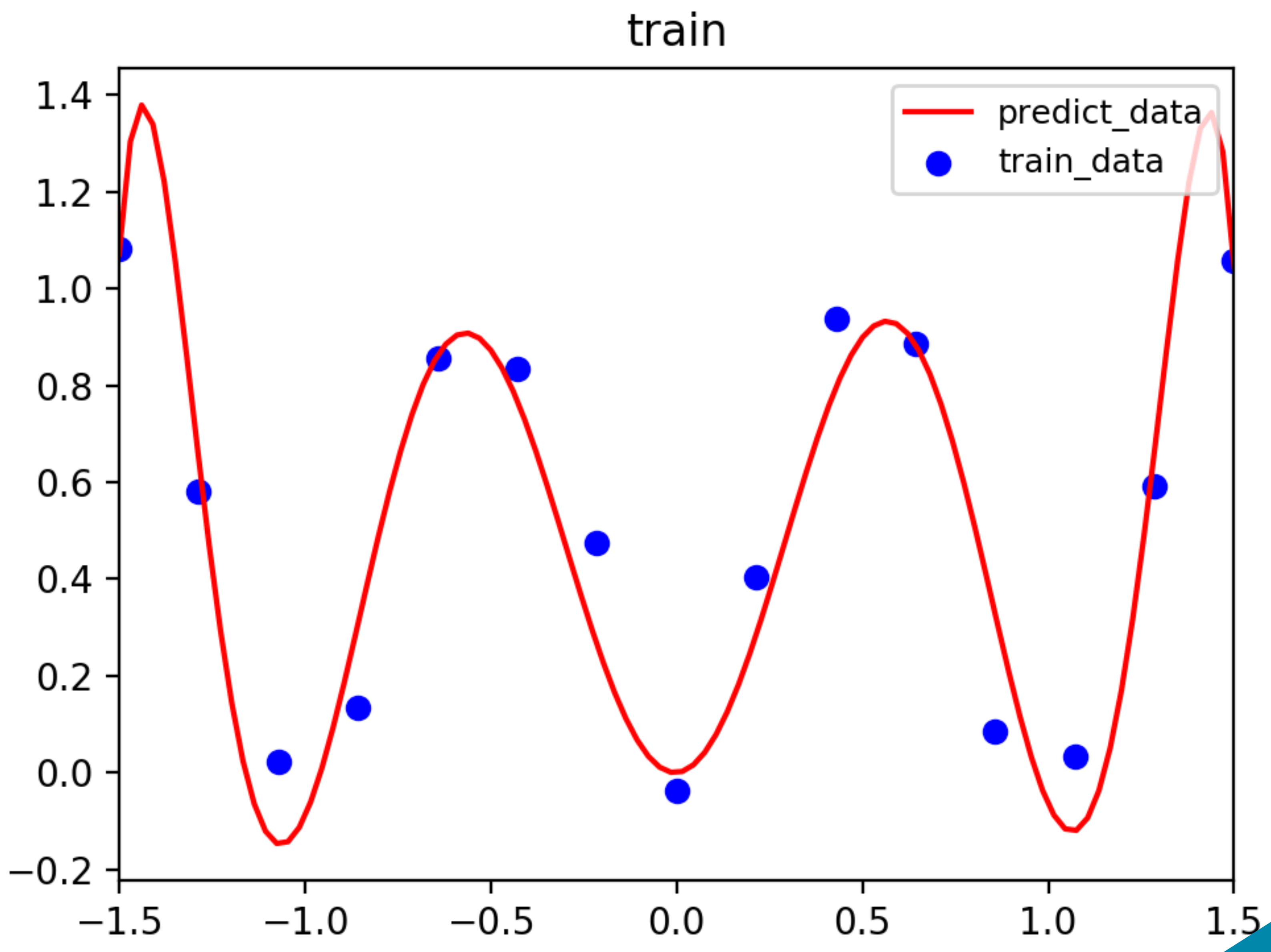
実験結果(最小二乗法，最尤推定)

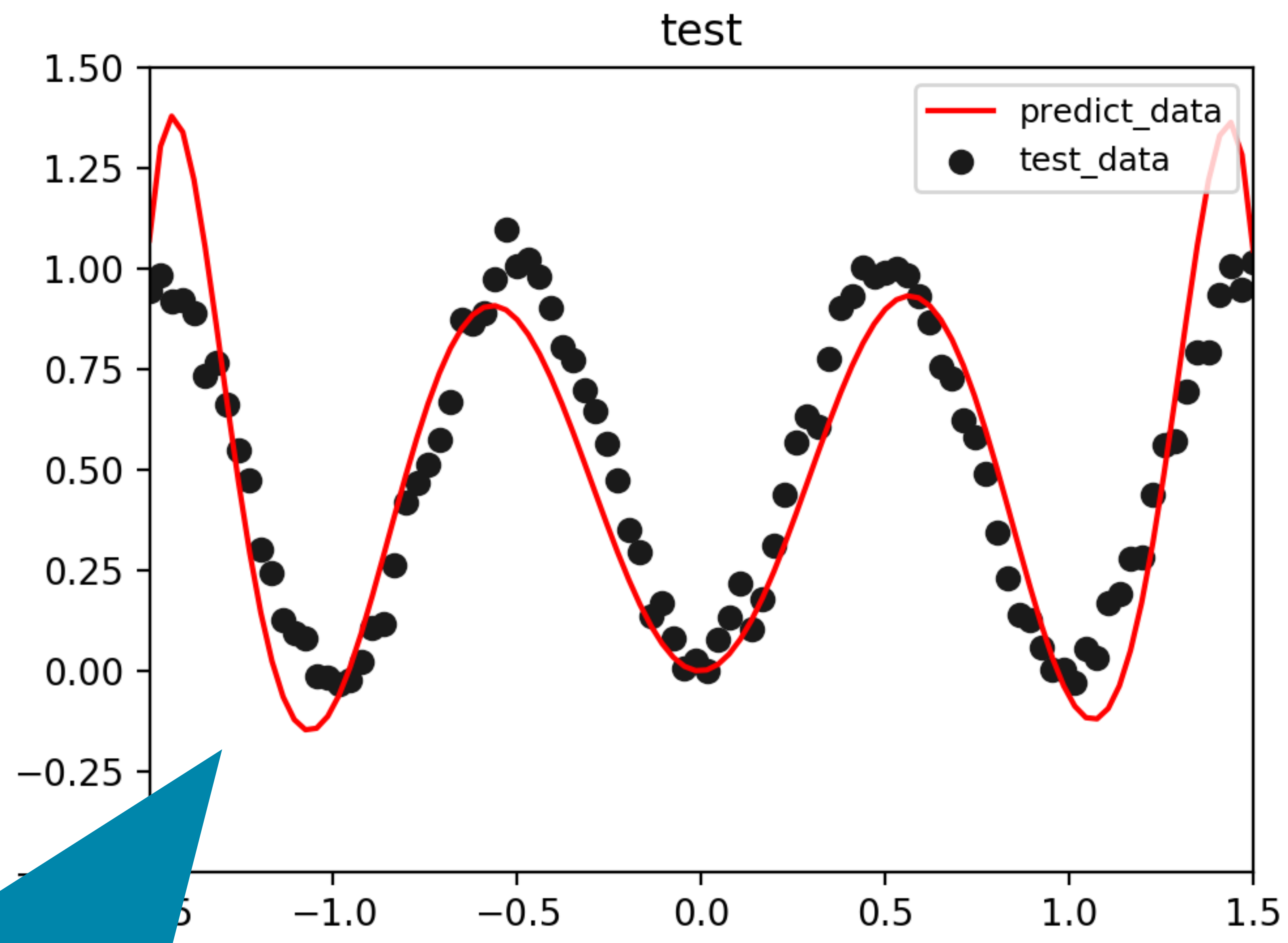
訓練データが少ない部分の
予測精度は下がる.





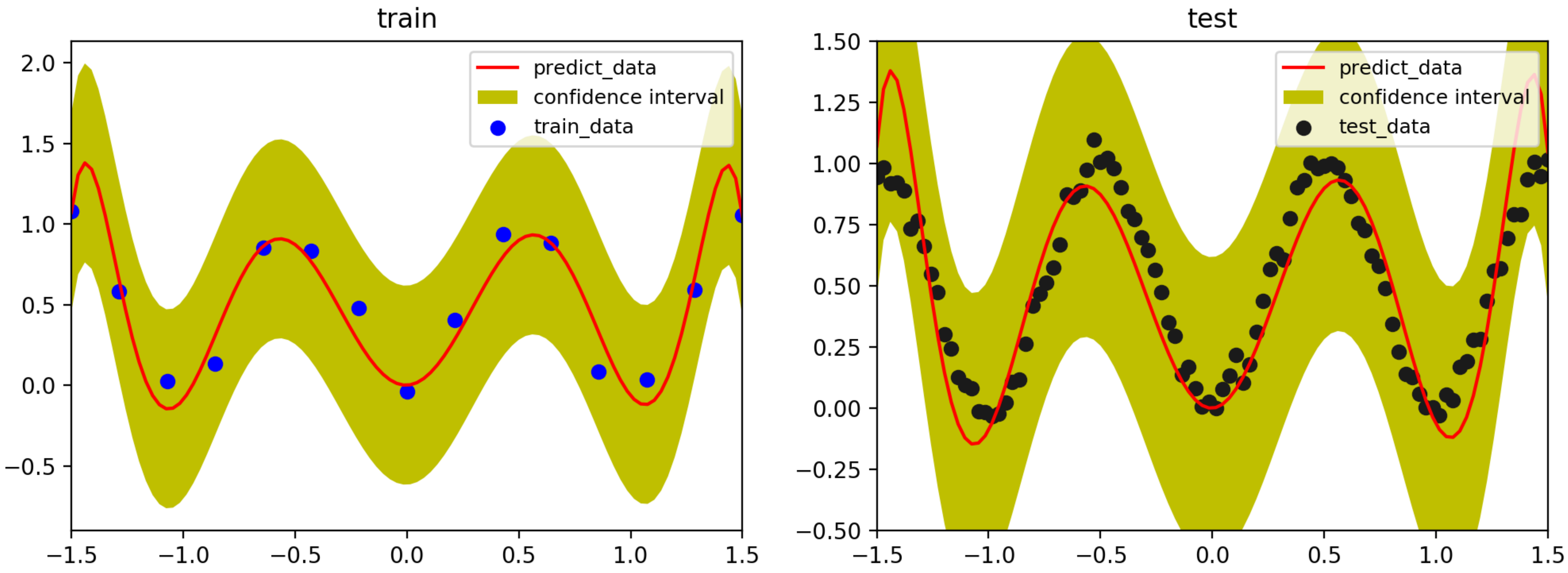
MAP推定($\alpha = 10, \beta = 10$)





大きく外れる部分が減少した.

ベイズ推定($\alpha = 10, \beta = 10$)

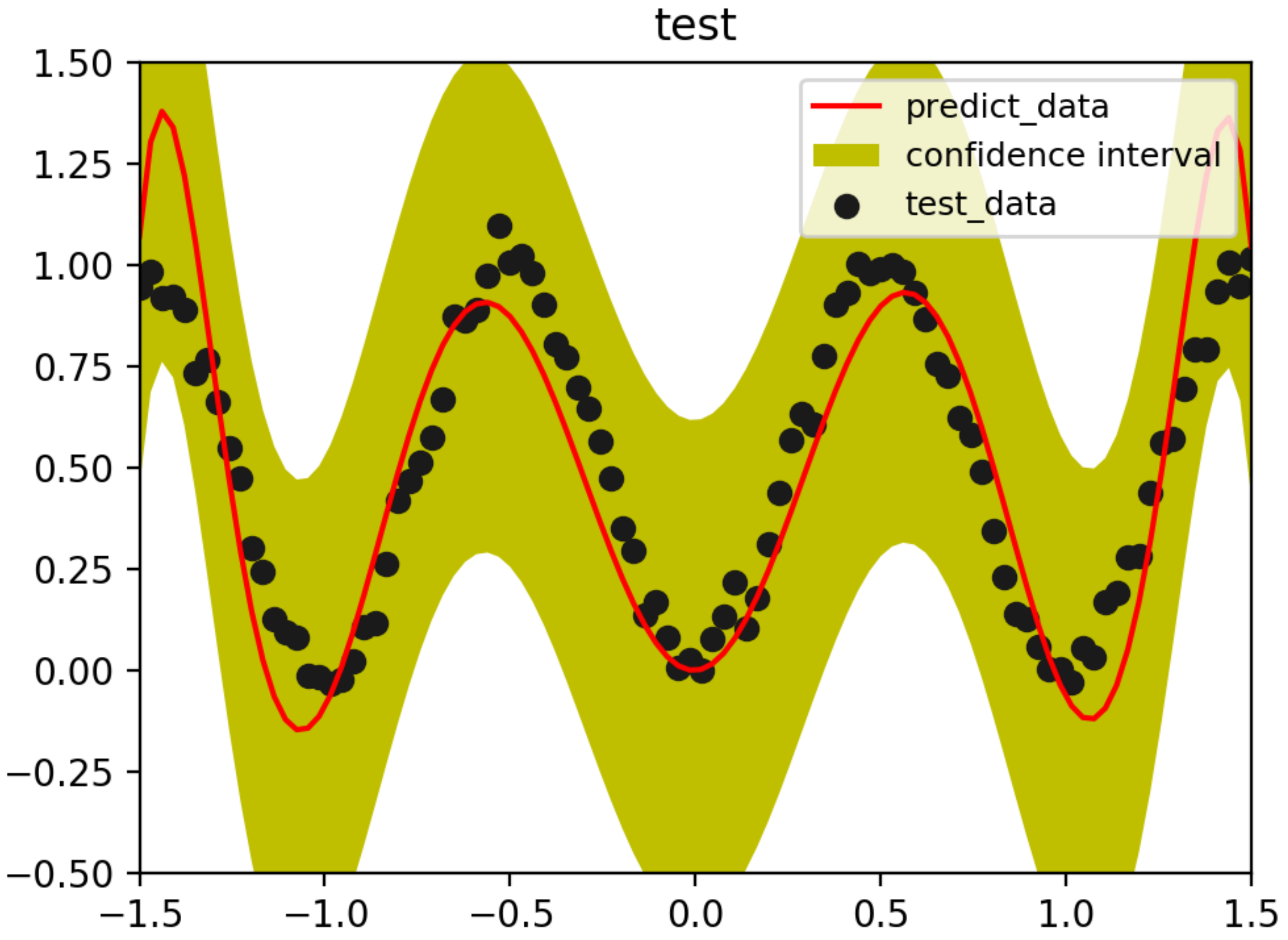


決定係数R^2

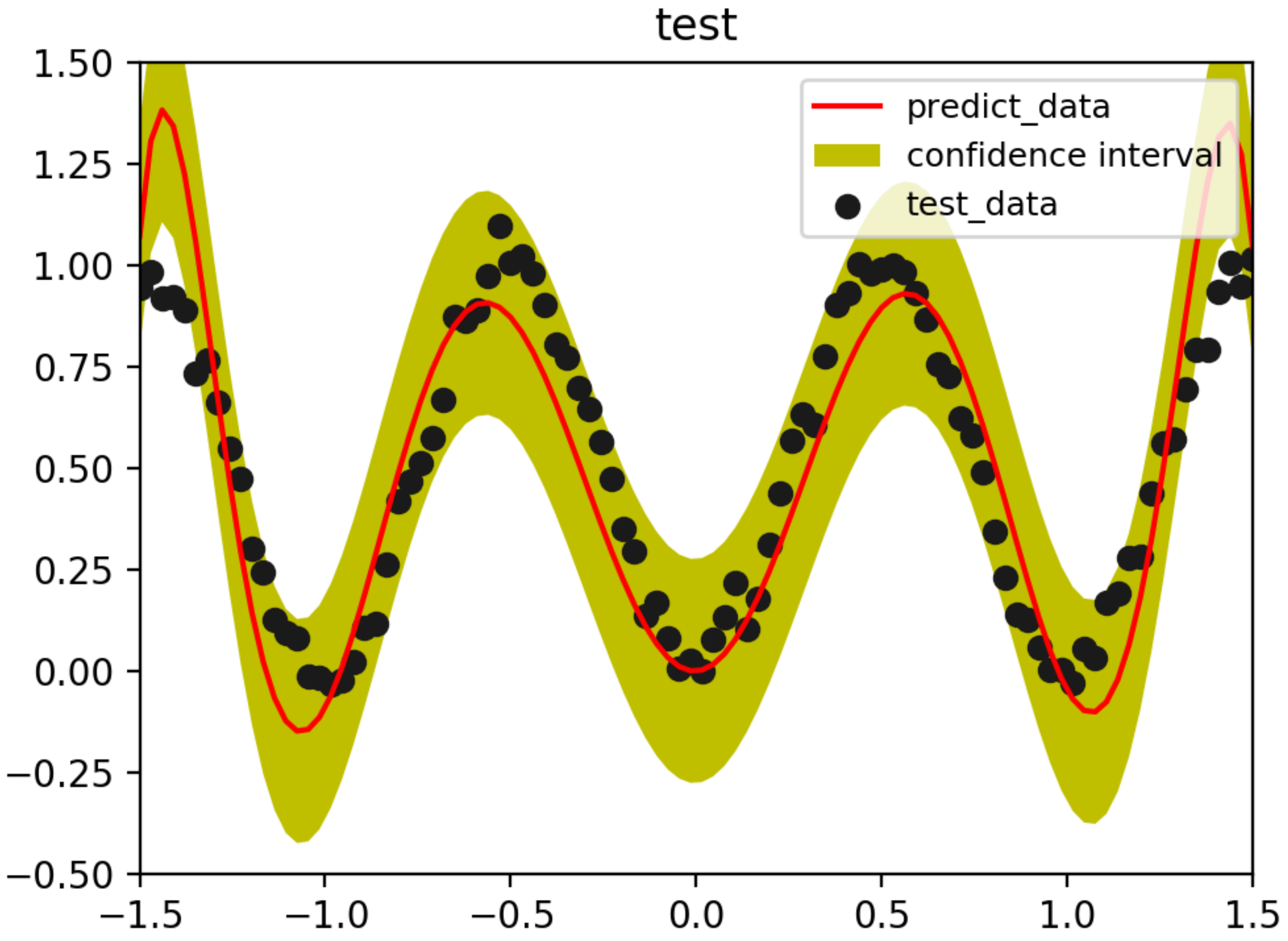
$$R^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (11)$$

手法	最小二乗推定，最尤推定	MAP推定
R^2	0.760122979	0.769478635

ベイズ推定($\beta = 10,50,100,1000$)

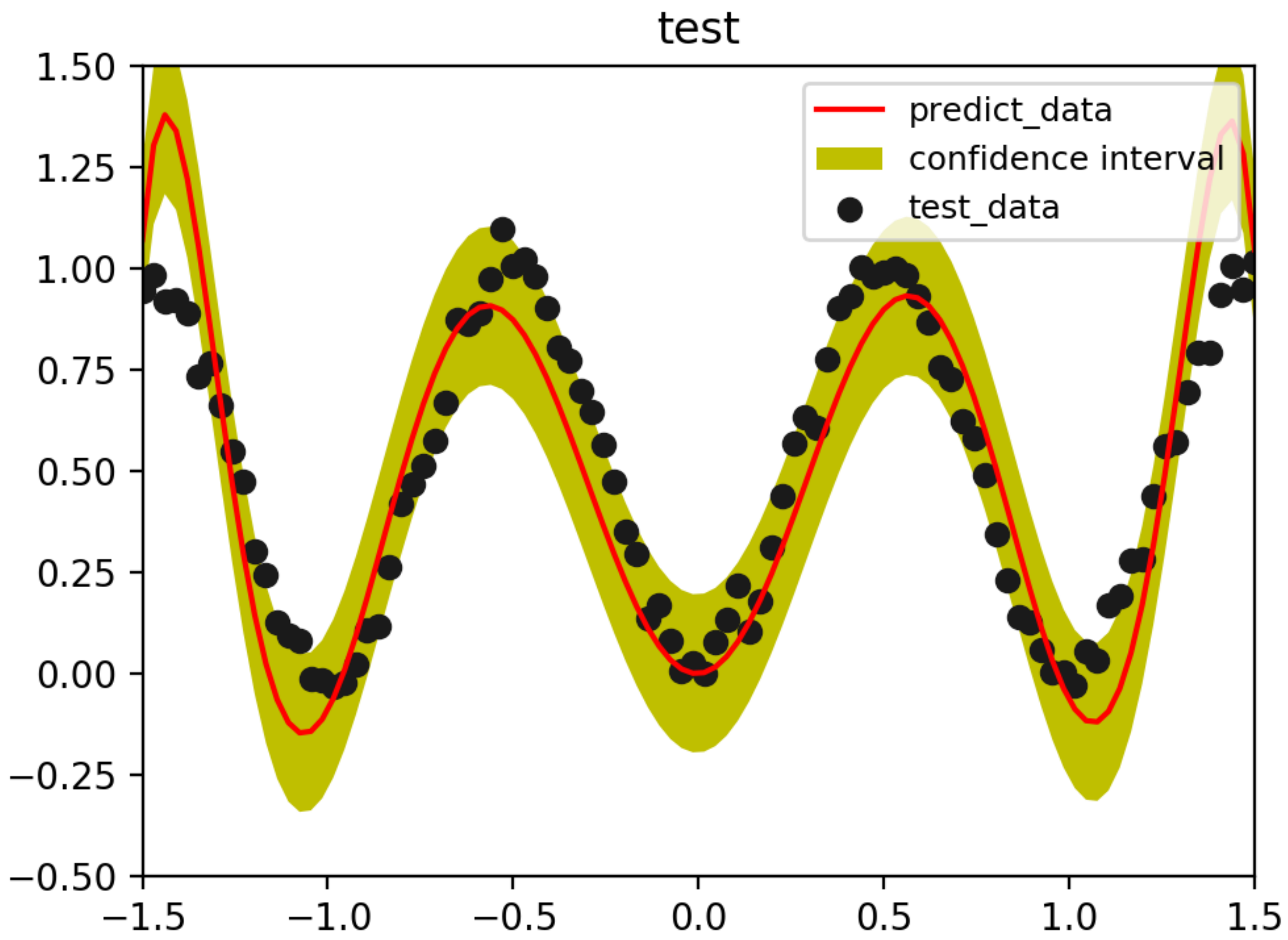


$\beta = 10$

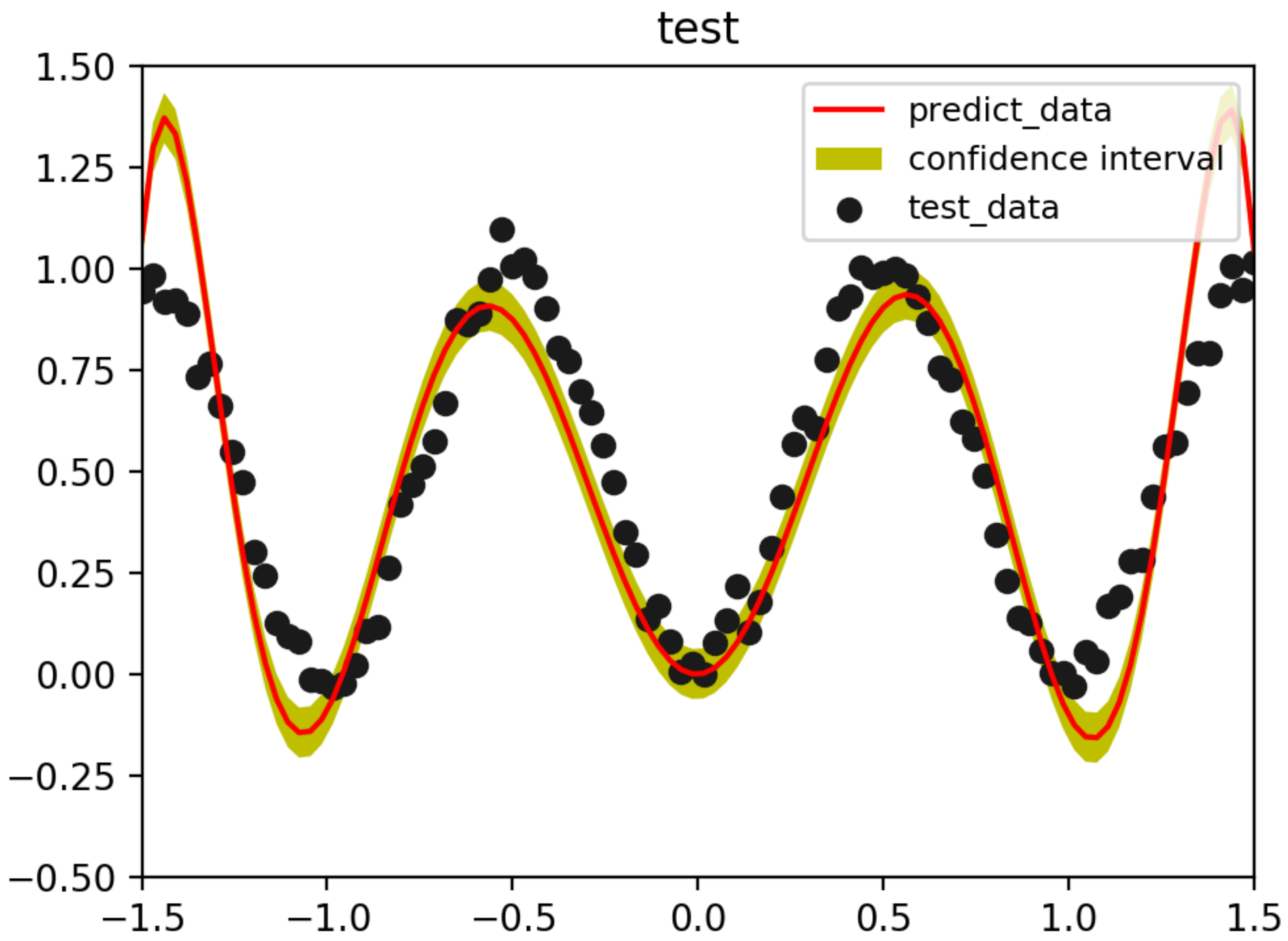


$\beta = 50$

ベイズ推定($\beta = 10,50,100,1000$)



$\beta = 100$



$\beta = 1000$

MAP推定($\beta = 10,50,100,1000$)のR^2

β	10	50	100	1000
R^2	0.76947	0.77227	0.769478	0.761849

誤差が小さくなっている.

過学習を起こしている.

まとめ

まとめ

事前に与えられる情報に応じて，適切に推定方法を変えることが重要であることがわかった．

訓練データが少ない部分について，誤差を小さくすることができなかったことが今後の課題である．