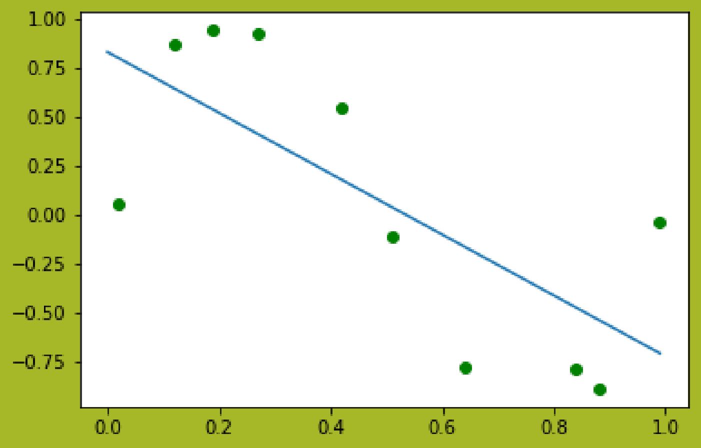
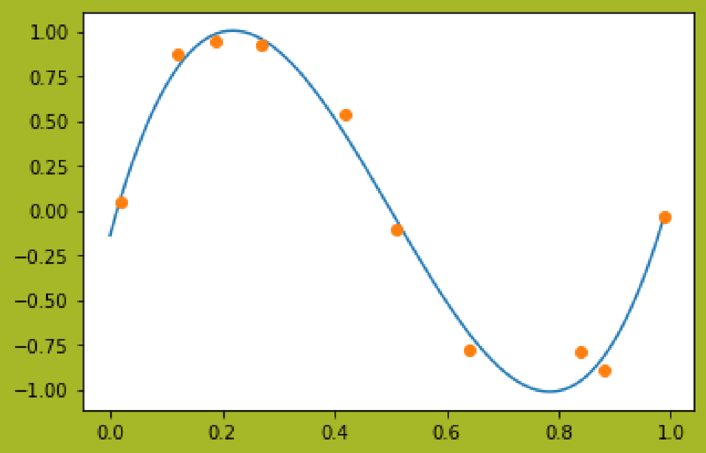
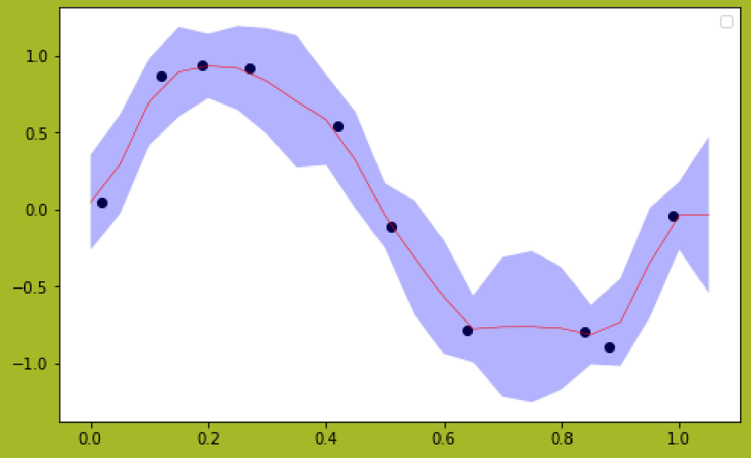
# 線形回帰とガウス過程回帰

中央大学理工学部物理学科　学部４年
香取研究室　石井理矩斗



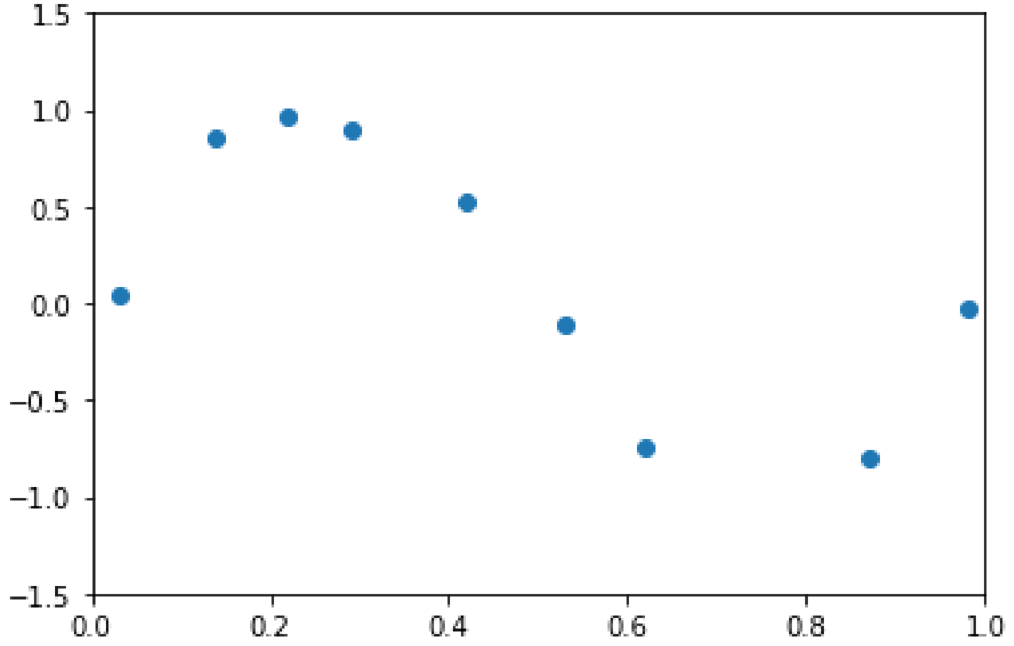



# 回帰とは？

- Yが連続値のとき、データにY=f(X)という定量的な関係をあてはめること。

- Xが一次元ならば単回帰、二次元以上ならば重回帰という。

# 今回扱うデータ




$X = [0.03, 0.14, 0.22, 0.29, 0.42, 0.53, 0.62, 0.87, 0.98]$

$Y = [0.04, 0.85, 0.96, 0.90, 0.53, -0.11, -0.75, -0.80, -0.03]$

一次関数?　　三次関数?　　三角関数?

# 単回帰(最小二乗法)

- $y = a + bx \ \ (a, b \in \mathbb{R})$と全データとの誤差の総和が最小となる$a, b$を求める。
- 誤差は負になることもあるので、二乗したものの総和を最小にする。

$$E = \sum_{n=1}^{N} \left(y_n - (a + bx_n)\right)^2$$

# 単回帰(最小二乗法)続き

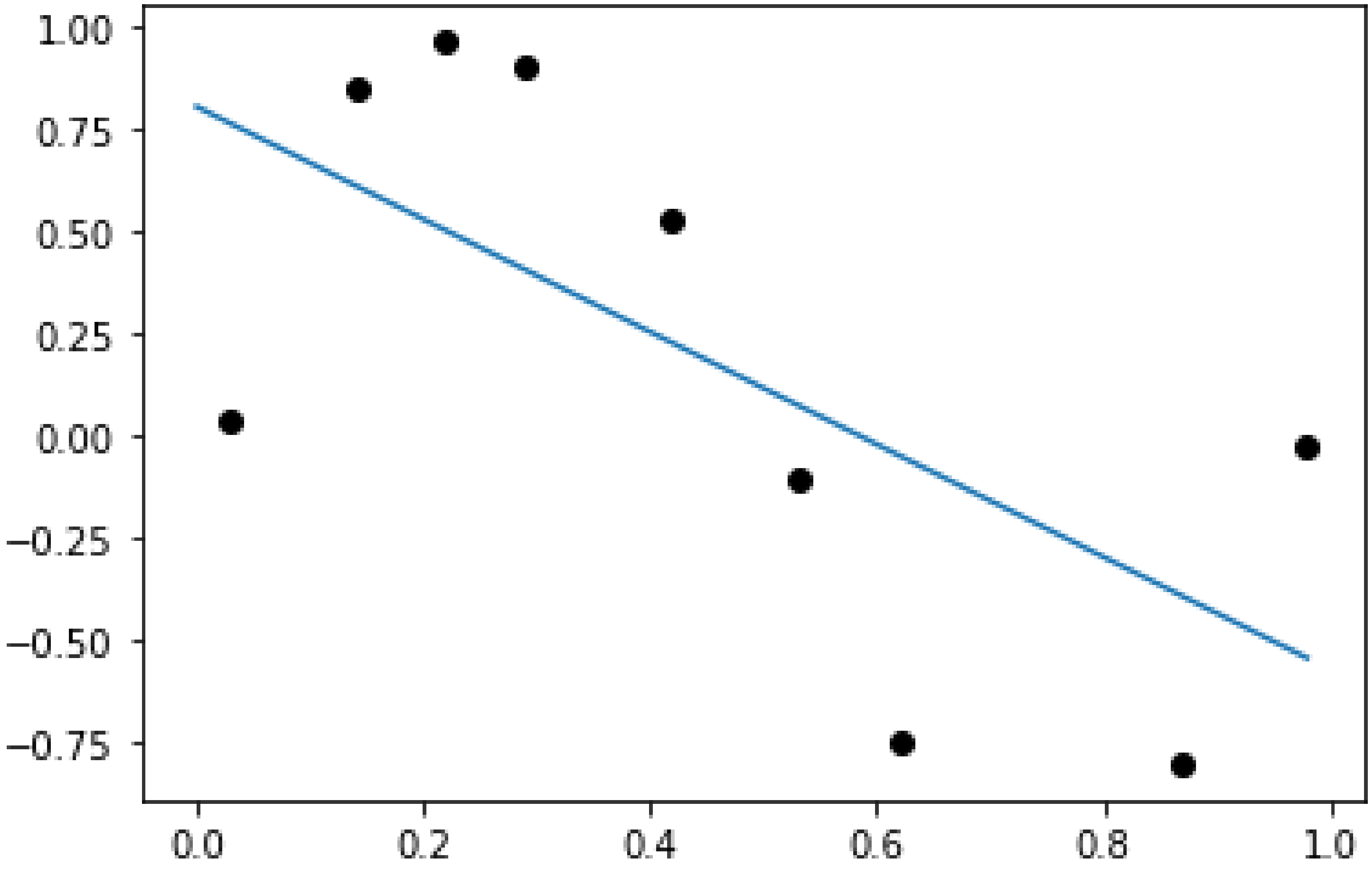
$\frac{\partial E}{\partial a} = 0, \frac{\partial E}{\partial b} = 0$として、

$$\begin{cases} aN + b\sum_{n=1}^{N} x_n = \sum_{n=1}^{N} y_n \\ a\sum_{n=1}^{N} x_n + b\sum_{n=1}^{N} x_n^2 = \sum_{n=1}^{N} x_n y_n \end{cases}$$

これは$a, b$についての連立一次方程式なので、簡単に解ける。

$$\binom{a}{b} = \frac{1}{N\sum_n x_n^2 - (\sum_n x_n)^2} \begin{pmatrix} \sum_n x_n^2 \sum_n y_n - \sum_n x_n \sum_n x_n y_n \\ N\sum_n x_n y_n - \sum_n x_n \sum_n y_n \end{pmatrix}$$

$$y = 0.802 - 1.37x$$



# 線形回帰モデル

- 単回帰よりも複雑な関係を表すには、$y = w_0 + w_1 x + w_2 x$ のような二次式や、さらに三角関数を加えた $y = w_0 + w_1 x + w_2 x + w_3 \sin x$ のような関数が考えられる。
- 一般に、$x$の関数 $\phi_1(x), \cdots \phi_H(x)$ を$H$個用意することで、入出力関係の表現力を大幅に上げることができる。

$$\boldsymbol{w} = \begin{pmatrix} w_0 \\ \vdots \\ w_H \end{pmatrix}, \boldsymbol{y} = \begin{pmatrix} y_1 \\ \vdots \\ \vdots \\ y_N \end{pmatrix}, \boldsymbol{\Phi} = \begin{pmatrix} 1 & \phi_1(x_1) & \cdots & \phi_H(x_1) \\ 1 & \phi_1(x_2) & \cdots & \phi_H(x_2) \\ \vdots & & & \vdots \\ 1 & \phi_1(x_N) & \cdots & \phi_H(x_N) \end{pmatrix}$$

$$E = \sum_{n=1}^{N} \left(y_n - \boldsymbol{w}^T \boldsymbol{x}_n\right)^2 = \boldsymbol{y}^T \boldsymbol{y} - 2\boldsymbol{w}^T\left(\boldsymbol{\Phi}^T \boldsymbol{y}\right) + \boldsymbol{w}^T \boldsymbol{\Phi}^T \boldsymbol{\Phi} \boldsymbol{w}$$

# 線形回帰モデル(続き)

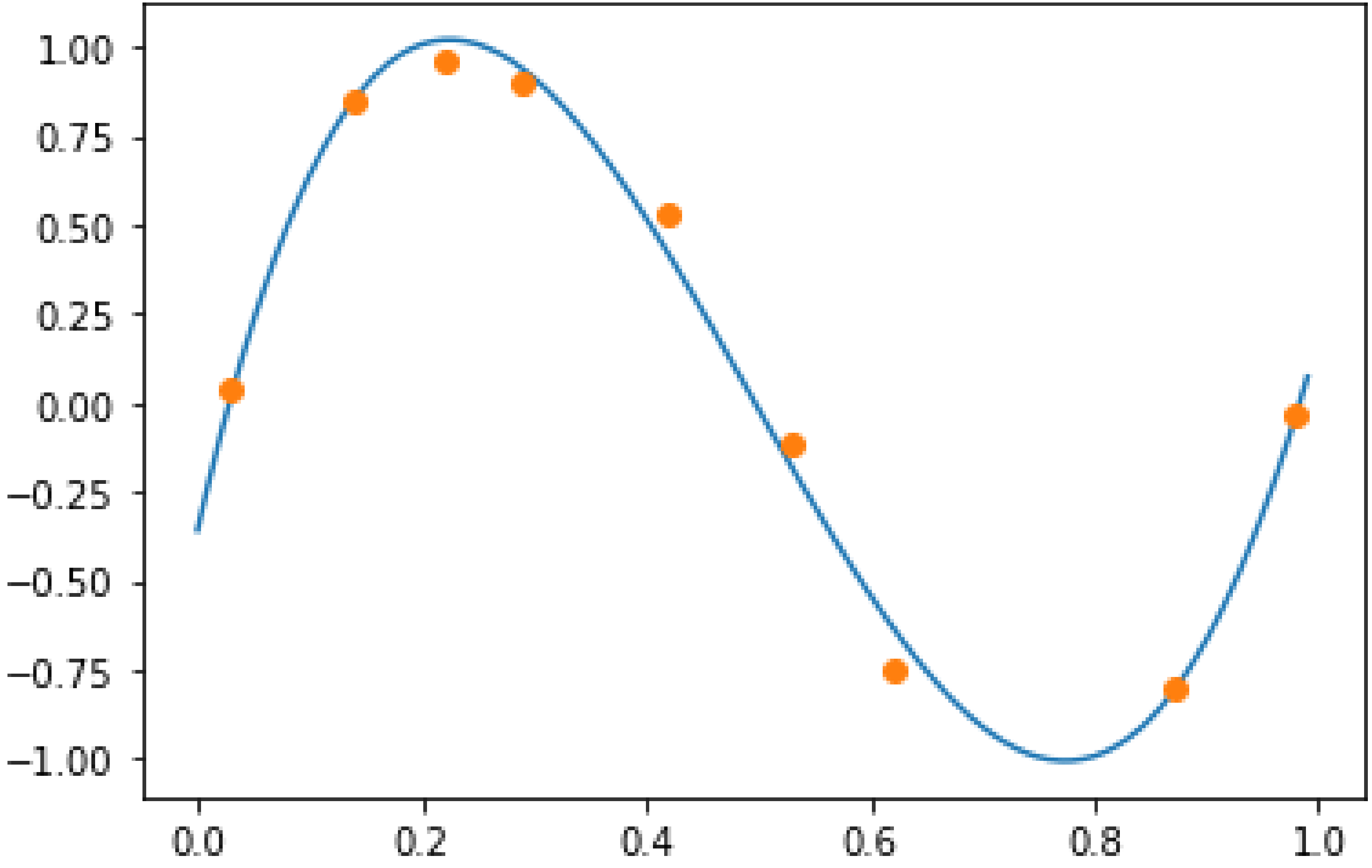
$$\frac{\partial E}{\partial \boldsymbol{w}} = -2\Phi^T \boldsymbol{y} + 2\Phi^T \Phi \boldsymbol{w} = 0$$

$$\boldsymbol{w} = \left(\boldsymbol{\Phi}^T \boldsymbol{\Phi}\right)^{-1} \boldsymbol{\Phi}^T \boldsymbol{y}$$

ただし、$\left(\boldsymbol{\Phi}^T \boldsymbol{\Phi}\right)$の逆行列が存在する必要がある。(今考えているデータはこれを満たす。)
満たさない場合はデータ間に定数倍の関係があるとき
→定数倍の連立方程式を解くようなこと

- 今回は、特徴ベクトル $\boldsymbol{\phi}(\boldsymbol{x}) = \left(1, x, x^2, \sin x\right)^T$を用いる。

$$y = -0.358 + 182x - 39.8x^2 - 168\sin x$$



# ガウス過程回帰

- 今までは基底関数を明示的に与えていたが、ガウス分布の形をした基底関数を導入し、それぞれ適当に重み付け、一定間隔に配置することで、ほとんど任意の形の関数を表せる。これを動径基底関数回帰という。

$$\phi_h(x) = \exp\left(-\frac{(x-\mu_h)^2}{\sigma^2}\right)$$

$$y = \sum_{h=-H}^{H} w_h \exp\left(-\frac{(x-\mu_h)^2}{\sigma^2}\right)$$

- この方法では、基底関数を20個配置したとき、$w$は20個、$x$を二次元にすると$w$は$20^2$個のように指数的に増え扱えなくなる。これを次元の呪いという。

# ガウス過程回帰(続き)

- $D$次元のベクトル$\boldsymbol{x} = (x_1, \cdots x_D)^T$が平均$\boldsymbol{\mu} = (\mu_1, \cdots \mu_D)^T$、共分散行列$\boldsymbol{\Sigma}$のガウス分布に従っているとき、多変量ガウス分布の確率密度関数は以下で与えられる。

$$\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^D\sqrt{|\boldsymbol{\Sigma}|}}\exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right)$$

- $\boldsymbol{\Sigma}$の共分散行列は、その(i, j)成分が$x_i$と$x_j$の共分散を表している行列。

$$\boldsymbol{\Sigma} = \mathbb{E}[\boldsymbol{x}\boldsymbol{x}^T] - \mathbb{E}[\boldsymbol{x}]\mathbb{E}[\boldsymbol{x}]^T$$

# ガウス過程回帰(続き)

- 線形回帰モデル $\boldsymbol{y} = \boldsymbol{\Phi w}$ に$\boldsymbol{w} \sim \mathcal{N}(\boldsymbol{w}|\boldsymbol{0}, \lambda^2 \boldsymbol{I})$を仮定すると、$\boldsymbol{\Phi}$は定数行列→ $\boldsymbol{\Phi w}$は線形変換→ $\mathrm{p}(\boldsymbol{y})$もガウス分布、ということが分かる。
- さらにこれに期待値をとったものは、

$$\mathbb{E}[\boldsymbol{y}] = \mathbb{E}[\boldsymbol{\Phi w}] = 0$$

$$\Sigma = \mathbb{E}[\boldsymbol{y}\boldsymbol{y}^T] - \mathbb{E}[\boldsymbol{y}]\mathbb{E}[\boldsymbol{y}]^T = \lambda^2 \boldsymbol{\Phi}\boldsymbol{\Phi}^T$$

  つまり、

$$\boldsymbol{y} \sim \mathcal{N}\left(\boldsymbol{0}, \lambda^2 \boldsymbol{\Phi}\boldsymbol{\Phi}^T\right) = \mathcal{N}(\boldsymbol{0}, \boldsymbol{K})$$

  ただし、 $\boldsymbol{K}$はカーネル行列という。

# ガウス過程回帰(続き)

- ガウス過程の意味

$$K_{nn'} = \lambda^2 \boldsymbol{\phi}(\boldsymbol{x}_n)^T \boldsymbol{\phi}(\boldsymbol{x}_{n'})$$

共分散が大きいときは似た値をとりやすいということだから、特徴ベクトルの内積が大きいとき、対応する出力値も似た値をとるということ。
ガウス過程とは、この関係を数学的に表現するための道具。

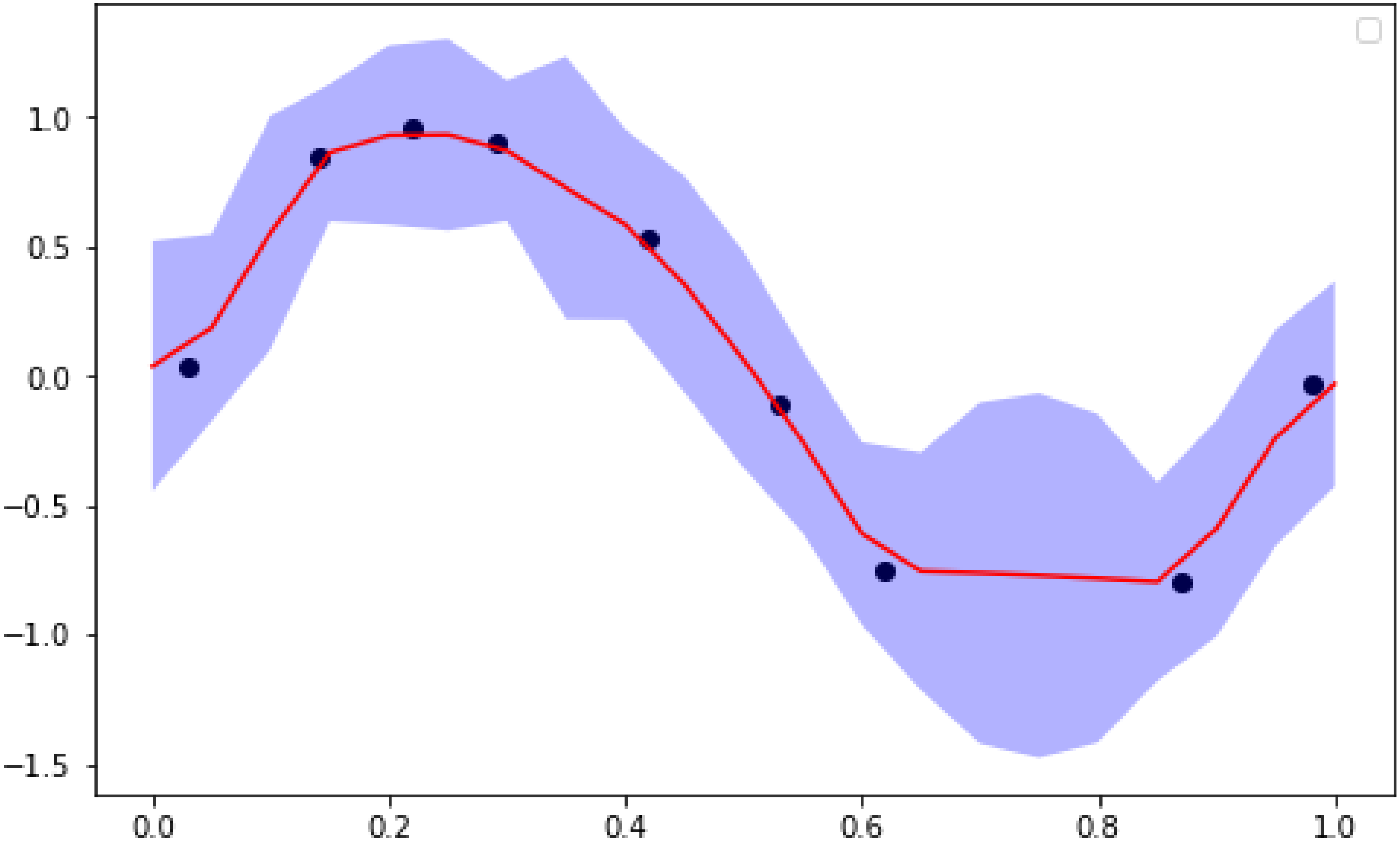
- カーネルトリック

$$k(\boldsymbol{x}_n, \boldsymbol{x}_{n'}) = \boldsymbol{\phi}(\boldsymbol{x}_n)^T \boldsymbol{\phi}(\boldsymbol{x}_{n'})$$

カーネル行列はカーネル関数$k(\boldsymbol{x}_n, \boldsymbol{x}_{n'})$から求まるので、特徴ベクトル
$\boldsymbol{\phi}(\boldsymbol{x}_n)$を明示的に表現することなしに、カーネル行列を得ることができる。
これをカーネルトリックという。

# ガウス過程回帰(続き)

- ガウス過程回帰の予測分布
  データに含まれない$x^*$での$y^*$の値を予測する際、ガウス過程の場合はその
  データをを含めた全体がガウス過程に従う。すなわち、
  $$\begin{pmatrix} \boldsymbol{y} \\ y^* \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} \boldsymbol{K} & \boldsymbol{k}_* \\ \boldsymbol{k}_*^T & k_{**} \end{pmatrix}\right)$$
  ただし、$\boldsymbol{k}_*^T = \left(k(\boldsymbol{x}^*, \boldsymbol{x}_1), k(\boldsymbol{x}^*, \boldsymbol{x}_2), \cdots, k(\boldsymbol{x}^*, \boldsymbol{x}_N)\right), k_{**} = k(x^*, x^*)$
  それぞれ*のついた文字はデータに含まれない予測したい点に依存している。
  $\boldsymbol{x}^*$と$D$が与えられたときの条件付き確率、
  ガウス過程の予測分布は以下で与えられる。
  $$p(y^* | \boldsymbol{x}^*, D) = \mathcal{N}\left(\boldsymbol{k}_*^T \boldsymbol{K}^{-1} \boldsymbol{y}, k_{**} - \boldsymbol{k}_*^T \boldsymbol{K}^{-1} \boldsymbol{k}_*\right)$$


1.0
0.5
0.0
−0.5
−1.0
−1.5
0.0
0.2
0.4
0.6
0.8
1.0

# ガウス過程回帰の計算コスト

- ガウス過程回帰はカーネル行列やその逆行列も求める計算コストが大きく、データが多いと扱いにくい。そこで、最後に計算コストを減らす工夫を紹介する。

- オーダー表示
  計算コストを表す表示法にオーダー表示がある。データ数が$N$、必要な演算数が$N(N-1)/2$のとき、オーダー表示は$O\left(N^2\right)$となる。なぜなら$N^2$の項は定数倍、$N$の項に比べ、$N$が大きくなった際に急速に大きくなるからである。

# ガウス過程回帰の計算コスト(続き)

- 補助変数法
  いくつかのデータ点をうまく代表する新たなデータ点として置きなおし、それを用いて計算する。
- ミニバッチと確率的勾配法
  データをミニバッチと呼ばれる100,500個程度のサイズに分け、それぞれの損失の平均をステップごとに計算し、パラメータ更新を進める。
- 格子状補助入力点配置
  補助入力点を格子状に配置することで、クロネッカー法、テプリッツ法、局所的カーネル補間の利点を取り入れる。

| 計算 | メモリ消費量オーダー | 演算量オーダー |
|---|---|---|
| 通常のガウス過程回帰 | $O(N^2)$ | $O(N^3)$ |
| 補助変数法 | $O(NM+M^2)$ | $O(NM^2+M^3)$ |
| 変分法+ミニバッチ | $O(M^2)$ | $O(M^3)$ |
| 格子型補助入力点配置 | $O(N+M^{1/2})$ | $O(N+M^{3/2})$ |

# まとめ

- 今回は一次元データを回帰することのみを考えていたが、ガウス過程回帰では特徴ベクトルを明示する必要がなく、重み$w$を積分消去することができるため高次元でも予測可能。
- ガウス過程回帰では線形回帰と異なり、その回帰結果の自信のある領域とない領域の違いが分かる。
- ガウス過程回帰では計算量の多さがボトルネックとなっているが、巧妙な工夫によって計算コストを大幅に節約できる。
- 機械学習や多くの分野においてもガウス過程の使用は急速に広まっている。

# 参考文献


- [1]持橋大地,大羽成征 . **ガウス過程と機械学習**. 機械学習プロフェッショナルシリーズ. 講談社,2019.
- [2] Pythonで「線形回帰」と"確率版の線形回帰"である「ベイズ線形回帰」 https://qiita.com/ysdyt/items/05a884354741bd9ca82b