機械学習 Machine Learning

線形回帰モデル Linear Models for Regression

福嶋 誠 Makoto Fukushima

情報科学部
School of Informatics and Data Science

線形回帰 Linear regression

訓練データ Training data

入力変数 Input variables: $\{\mathbf{x}_n\}$

目標変数 Target variables: $\{t_n\}$

$n = 1, \ldots, N$

新しいデータ New data $\mathbf{x} \longrightarrow y(\mathbf{x}, \mathbf{w}) \Rightarrow t$ 予測 Prediction

訓練データより学習可能なパラメータ
Parameter that can be learned from the training data $\mathbf{w}$

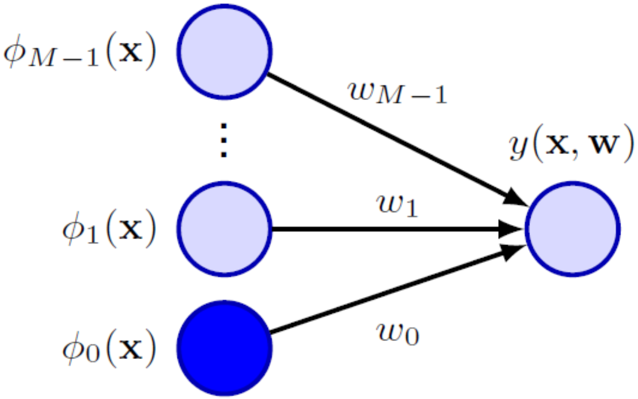
$w_0, \ldots, w_D$ に関する線形関数
A linear function of $w_0, \ldots, w_D$

最もシンプルな回帰モデル
The simplest model for regression

$$y(\mathbf{x}, \mathbf{w}) = w_0 + w_1 x_1 + \ldots + w_D x_D$$

$$\mathbf{x} = (x_1, \ldots, x_D)^{\mathrm{T}} \quad \mathbf{w} = (w_0, \ldots, w_D)^{\mathrm{T}}$$

線形回帰 Linear regression
基底関数 Basis functions



入力変数に関して非線形な関数の線形結合
Linear combinations of fixed nonlinear functions of the input variables

$$y(\mathbf{x}, \mathbf{w}) = w_0 + \sum_{j=1}^{M-1} w_j \phi_j(\mathbf{x})$$

$\phi_j(\mathbf{x})$ 基底関数 Basis functions

$$\downarrow$$

$$y(\mathbf{x}, \mathbf{w}) = \sum_{j=0}^{M-1} w_j \phi_j(\mathbf{x}) = \mathbf{w}^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x})$$

$\phi_0(\mathbf{x}) = 1$ w_0 バイアス Bias

$\mathbf{w} = (w_0, \ldots, w_{M-1})^{\mathrm{T}}$ $\boldsymbol{\phi} = (\phi_0, \ldots, \phi_{M-1})^{\mathrm{T}}$

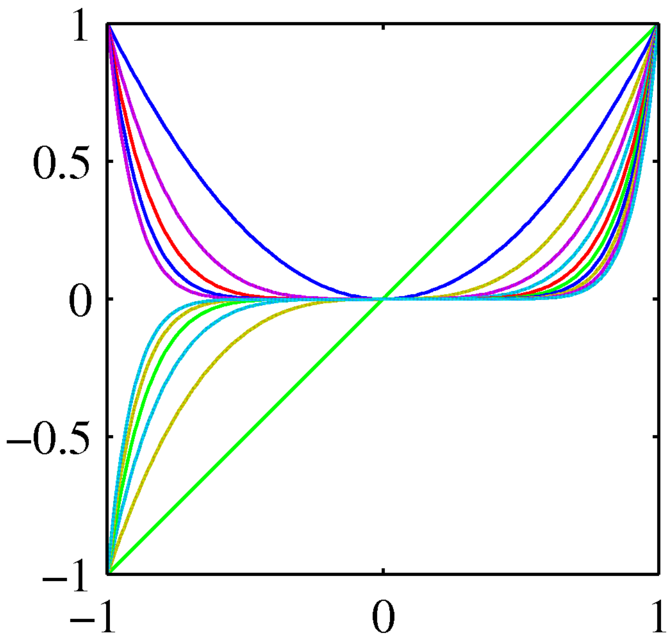
線形回帰 Linear regression
基底関数 Basis functions

$$y(\mathbf{x}, \mathbf{w}) = \sum_{j=0}^{M-1} w_j \phi_j(\mathbf{x}) = \mathbf{w}^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x})$$

基底関数の例
An example of the basis functions

多項式 Polynomial functions

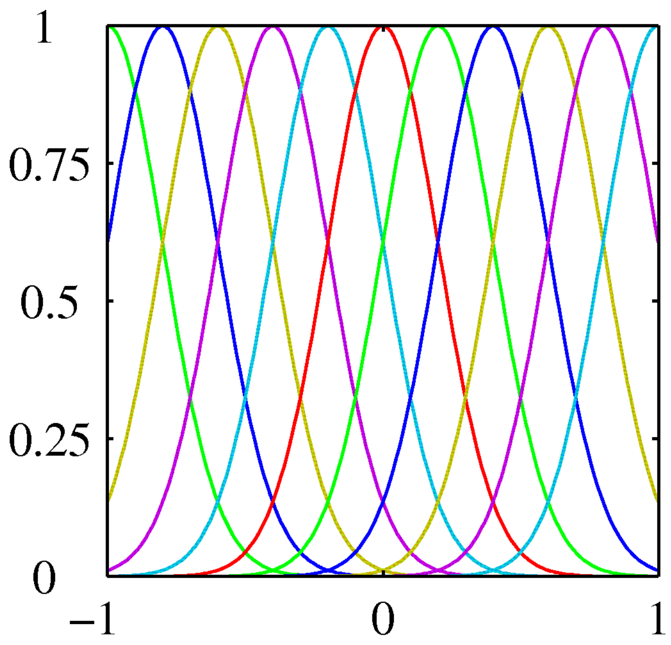
$$\phi_j(x) = x^j.$$



線形回帰 Linear regression
基底関数 Basis functions

$$y(\mathbf{x}, \mathbf{w}) = \sum_{j=0}^{M-1} w_j \phi_j(\mathbf{x}) = \mathbf{w}^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x})$$

基底関数の例
An example of the basis functions



ガウス基底関数 Gaussian basis functions

$$\phi_j(x) = \exp\left\{-\frac{(x-\mu_j)^2}{2s^2}\right\}$$

線形回帰 Linear regression
基底関数 Basis functions

$$y(\mathbf{x}, \mathbf{w}) = \sum_{j=0}^{M-1} w_j \phi_j(\mathbf{x}) = \mathbf{w}^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x})$$

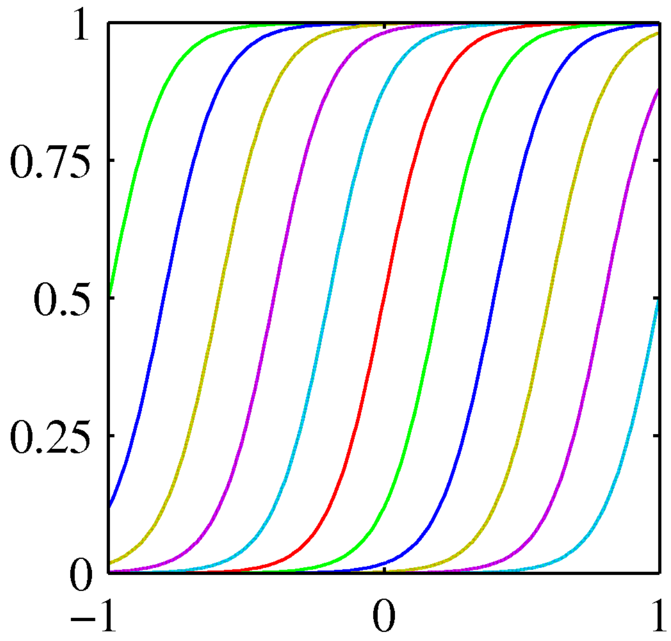
基底関数の例
An example of the basis functions

シグモイド基底関数 Sigmoid basis functions

$$\phi_j(x) = \sigma\left(\frac{x - \mu_j}{s}\right)$$

ロジスティックシグモイド関数
Logistic sigmoid function

$$\sigma(a) = \frac{1}{1 + \exp(-a)}.$$



線形回帰 Linear regression
尤度関数 Likelihood function

目標変数 t が決定論的な関数 $y(\mathbf{x}, \mathbf{w})$ と加法性のガウスノイズの和で与えられるとき
We assume that the target variable t is given by a deterministic function $y(\mathbf{x}, \mathbf{w})$ with additive Gaussian noise:

$$t = y(\mathbf{x}, \mathbf{w}) + \epsilon$$ ただし where $p(\epsilon|\sigma^2) = \mathcal{N}(\epsilon|0, \sigma^2)$

$$\longrightarrow \quad p(t|\mathbf{x}, \mathbf{w}, \sigma^2) = \mathcal{N}(t|y(\mathbf{x}, \mathbf{w}), \sigma^2)$$

入力 Inputs: $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ 目標値 Target values: $\mathbf{t} = [t_1, \ldots, t_N]^{\mathrm{T}}$

尤度関数
Likelihood function

$$p(\mathbf{t}|\mathbf{X}, \mathbf{w}, \sigma^2) = \prod_{n=1}^{N} \mathcal{N}(t_n|\mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n), \sigma^2)$$

線形回帰 Linear regression
尤度関数 Likelihood function

尤度関数の対数をとると
Taking the logarithm, we get

$$\begin{aligned}\ln p(\mathbf{t}|\mathbf{X}, \mathbf{w}, \sigma^2) &= \sum_{n=1}^{N} \ln \mathcal{N}(t_n|\mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n), \sigma^2) \\ &= -\frac{N}{2}\ln\sigma^2 - \frac{N}{2}\ln(2\pi) - \frac{1}{\sigma^2}E_D(\mathbf{w})\end{aligned}$$

二乗和誤差関数
Sum-of-squares error

$$E_D(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{t_n - \mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n)\}^2$$

線形回帰 Linear regression
最尤推定 Maximum likelihood

$$\nabla_{\mathbf{w}} \ln p(\mathbf{t}|\mathbf{X}, \mathbf{w}, \sigma^2) = \frac{1}{\sigma^2} \sum_{n=1}^{N} \left\{ t_n - \mathbf{w}^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x}_n) \right\} \boldsymbol{\phi}(\mathbf{x}_n)^{\mathrm{T}}$$

対数尤度関数の $\mathbf{w}$ についての勾配を0とおけば
Setting the gradient of the log likelihood function with respect to $\mathbf{w}$ to zero gives

$$0 = \sum_{n=1}^{N} t_n \boldsymbol{\phi}(\mathbf{x}_n)^{\mathrm{T}} - \mathbf{w}^{\mathrm{T}} \left(\sum_{n=1}^{N} \boldsymbol{\phi}(\mathbf{x}_n) \boldsymbol{\phi}(\mathbf{x}_n)^{\mathrm{T}} \right)$$

$$\Rightarrow \quad \mathbf{w}_{\mathrm{ML}} = \underbrace{\left(\boldsymbol{\Phi}^{\mathrm{T}} \boldsymbol{\Phi} \right)^{-1} \boldsymbol{\Phi}^{\mathrm{T}}}_{\rightarrow\ \boldsymbol{\Phi}^{\dagger}} \mathbf{t}$$

ムーア–ペンローズの擬似逆行列
Moore–Penrose pseudo-inverse matrix

ただし
where

$$\mathbf{t} = [t_1, \ldots, t_N]^{\mathrm{T}}$$

$$\boldsymbol{\Phi} = \begin{pmatrix} \phi_0(\mathbf{x}_1) & \phi_1(\mathbf{x}_1) & \cdots & \phi_{M-1}(\mathbf{x}_1) \\ \phi_0(\mathbf{x}_2) & \phi_1(\mathbf{x}_2) & \cdots & \phi_{M-1}(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_0(\mathbf{x}_N) & \phi_1(\mathbf{x}_N) & \cdots & \phi_{M-1}(\mathbf{x}_N) \end{pmatrix}$$

線形回帰 Linear regression
最尤推定 Maximum likelihood

$$\ln p(\mathbf{t}|\mathbf{X}, \mathbf{w}, \sigma^2) = \sum_{n=1}^{N} \ln \mathcal{N}(t_n|\mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n), \sigma^2)$$
$$= -\frac{N}{2}\ln\sigma^2 - \frac{N}{2}\ln(2\pi) - \frac{1}{\sigma^2}E_D(\mathbf{w})$$

$$E_D(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{t_n - w_0 - \sum_{j=1}^{M-1} w_j\phi_j(\mathbf{x}_n)\}^2$$

w_0 について最大化すると

Maximizing with respect to w_0, we see that

$$w_0 = \bar{t} - \sum_{j=1}^{M-1} w_j\overline{\phi_j}$$

$$\bar{t} = \frac{1}{N}\sum_{n=1}^{N} t_n$$

$$\overline{\phi_j} = \frac{1}{N}\sum_{n=1}^{N}\phi_j(\mathbf{x}_n)$$

線形回帰 Linear regression
最尤推定 Maximum likelihood

$$\ln p(\mathbf{t}|\mathbf{X},\mathbf{w},\sigma^2) = \sum_{n=1}^{N} \ln \mathcal{N}(t_n|\mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n),\sigma^2)$$
$$= -\frac{N}{2}\ln\sigma^2 - \frac{N}{2}\ln(2\pi) - \frac{1}{\sigma^2}E_D(\mathbf{w})$$

$$E_D(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{t_n - \mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n)\}^2$$

σ^2について最大化すると

Maximizing with respect to σ^2, we see that

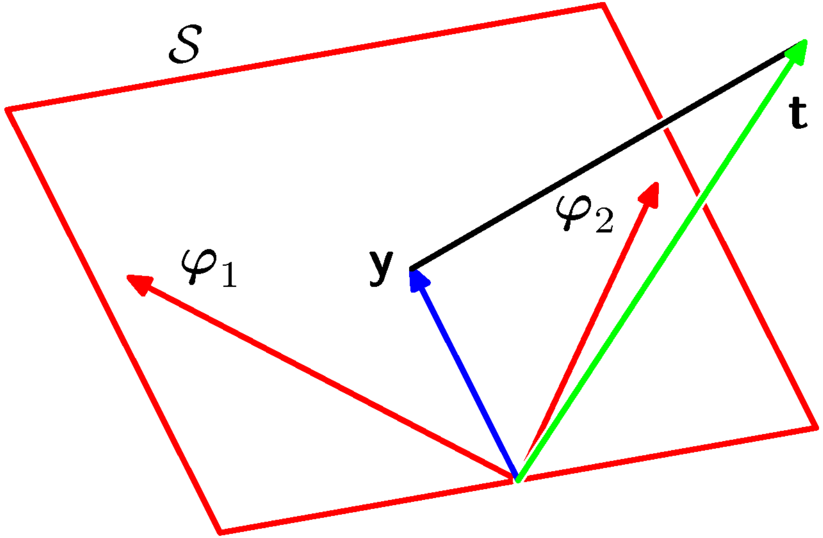
$$\sigma^2_{\mathrm{ML}} = \frac{1}{N}\sum_{n=1}^{N}\{t_n - \mathbf{w}^{\mathrm{T}}_{\mathrm{ML}}\boldsymbol{\phi}(\mathbf{x}_n)\}^2$$

線形回帰 Linear regression
最小二乗法の幾何学 Geometry of least squares

$$\mathbf{y} = \boldsymbol{\Phi}\mathbf{w}_{\mathrm{ML}} = [\boldsymbol{\varphi}_1, \ldots, \boldsymbol{\varphi}_M]\,\mathbf{w}_{\mathrm{ML}}.$$

$\mathbf{y} \in \mathcal{S} \subseteq \mathcal{T}$
M 次元
M dimensional

$\mathbf{t} \in \mathcal{T}$
N 次元
N dimensional



$\mathbf{w}_{\mathrm{ML}}$ は $\mathbf{t}$ と $\mathbf{y}$ の二乗ユークリッド距離を最小化する.

$\mathbf{w}_{\mathrm{ML}}$ minimizes the squared Euclidean distance between $\mathbf{t}$ and $\mathbf{y}$.

線形回帰 Linear regression
逐次学習 Sequential learning

逐次学習 Sequential learning

データ点を一度に1つだけ用いてモデルのパラメータを順次更新する.
Data points are considered one at a time and the model parameters are updated after each such presentation.

$$E = \sum_n E_n$$

$$E_n = \frac{1}{2}\left(t_n - \mathbf{w}^{(\tau)\mathrm{T}}\boldsymbol{\phi}_n\right)^2$$

確率的勾配降下法 Stochastic gradient descent

$$\begin{aligned}\mathbf{w}^{(\tau+1)} &= \mathbf{w}^{(\tau)} - \eta\nabla E_n \\ &= \mathbf{w}^{(\tau)} + \eta(t_n - \mathbf{w}^{(\tau)\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n))\boldsymbol{\phi}(\mathbf{x}_n).\end{aligned}$$

線形回帰 Linear regression
正則化最小二乗法 Regularized least squares

以下の誤差関数を考える Consider the error function:

$$E_D(\mathbf{w}) + \lambda E_W(\mathbf{w})$$

データに依存する誤差
Data-dependent error

正則化項
Regularization term

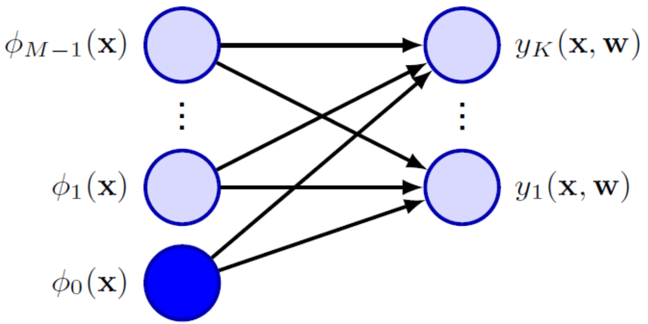
λ 正則化定数
Regularization coefficient

二乗和誤差関数と重みベクトルの二乗和を用いるときは
With the sum-of-squares error function and
the sum of the squares of the weight vector elements, we get

$$\frac{1}{2}\sum_{n=1}^{N}\{t_n - \mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n)\}^2 + \frac{\lambda}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w}$$

これは次のとき最小 which is minimized by $\mathbf{w} = \left(\lambda\mathbf{I} + \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}\right)^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{t}.$

線形回帰 Linear regression
出力変数が多次元の場合 Multiple outputs



$K > 1$ 次元の目標変数を予測したい場合
When we wish to predict $K > 1$ target variables:

$$p(\mathbf{t}|\mathbf{x}, \mathbf{W}, \sigma^2) = \mathcal{N}(\mathbf{t}|\mathbf{W}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}), \sigma^2\mathbf{I}) \qquad \mathbf{t} = (t_1, \ldots, t_K)^{\mathrm{T}}$$

$$\uparrow$$

$$\mathbf{y}(\mathbf{x}, \mathbf{w}) = \mathbf{W}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x})$$

入力 Inputs:

$$\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$$

目標値 Target values:

$$\mathbf{T} = [\mathbf{t}_1, \ldots, \mathbf{t}_N]^{\mathrm{T}}$$

対数尤度関数 Log likelihood function

$$\ln p(\mathbf{T}|\mathbf{X}, \mathbf{W}, \sigma^2) = \sum_{n=1}^{N} \ln \mathcal{N}(\mathbf{t}_n|\mathbf{W}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n), \sigma^2\mathbf{I})$$

$$= -\frac{NK}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{n=1}^{N}\left\|\mathbf{t}_n - \mathbf{W}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n)\right\|^2$$

線形回帰 Linear regression
出力変数が多次元の場合 Multiple outputs

尤度を最大化する $\mathbf{W}$ は
Maximizing the likelihood with respect to $\mathbf{W}$, we obtain

$$\mathbf{W}_{\mathrm{ML}} = \left(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}\right)^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{T}.$$

個々の目標変数 $\mathbf{t}_k$ について評価すれば
If we consider a single target variable $\mathbf{t}_k$, we see that

$$\mathbf{w}_k = \left(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}\right)^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{t}_k = \boldsymbol{\Phi}^{\dagger}\mathbf{t}_k \qquad \mathbf{t}_k = [t_{1k}, \ldots, t_{Nk}]^{\mathrm{T}}$$

決定理論 Decision theory

推論段階 Inference stage

$p(t|\mathbf{x})$ を求める.
Determine $p(t|\mathbf{x})$.

例 e.g. $p(t|\mathbf{x}, \mathbf{w}_{\mathrm{ML}}, \sigma^2_{\mathrm{ML}}) = \mathcal{N}(t|y(\mathbf{x}, \mathbf{w}_{\mathrm{ML}}), \sigma^2_{\mathrm{ML}})$

決定段階 Decision stage

ある $\mathbf{x}$ に対する t として最適な予測 $f(\mathbf{x})$ を行う.
For given $\mathbf{x}$, make optimal prediction $f(\mathbf{x})$ for t.

例 e.g. $y(\mathbf{x}, \mathbf{w}_{\mathrm{ML}})$

損失関数 Loss function $L(t, f(\mathbf{x}))$

例 e.g. $L(t, f(\mathbf{x})) = \{f(\mathbf{x}) - t\}^2$

期待損失 Expected loss $\mathbb{E}[L] = \iint L(t, f(\mathbf{x}))p(\mathbf{x}, t)\,\mathrm{d}\mathbf{x}\,\mathrm{d}t$

決定理論 Decision theory

$\mathbb{E}[L]$ を最小にする $f(\mathbf{x})$を選ぶ.

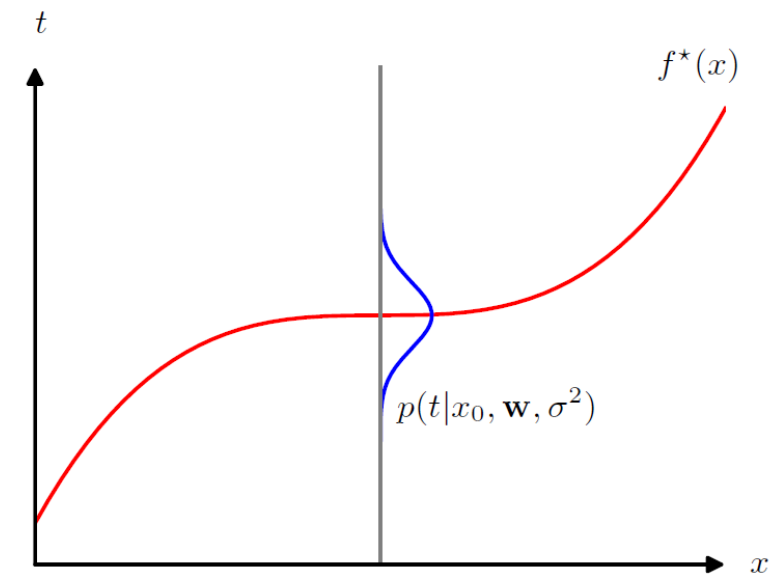
Choose $f(\mathbf{x})$ so as to minimize $\mathbb{E}[L]$.

$$\mathbb{E}[L] = \iint \{f(\mathbf{x}) - t\}^2 p(\mathbf{x}, t)\,\mathrm{d}\mathbf{x}\,\mathrm{d}t$$

変分法 Calculus of variations

$$\frac{\delta \mathbb{E}[L]}{\delta f(\mathbf{x})} = 2\int \{f(\mathbf{x}) - t\} p(\mathbf{x}, t)\,\mathrm{d}t = 0$$

$$\rightarrow \quad f^\star(\mathbf{x}) = \frac{1}{p(\mathbf{x})}\int t p(\mathbf{x}, t)\,\mathrm{d}t = \int t p(t|\mathbf{x})\,\mathrm{d}t = \mathbb{E}_t[t|\mathbf{x}]$$



$p(t|\mathbf{x}, \mathbf{w}, \sigma^2) = \mathcal{N}(t|y(\mathbf{x}, \mathbf{w}), \sigma^2)$ に対しては,

For $p(t|\mathbf{x}, \mathbf{w}, \sigma^2) = \mathcal{N}(t|y(\mathbf{x}, \mathbf{w}), \sigma^2)$,

$$\mathbb{E}[t|\mathbf{x}] = \int t p(t|\mathbf{x})\,\mathrm{d}t = y(\mathbf{x}, \mathbf{w})$$

決定理論 Decision theory

$$\mathbb{E}[L] = \iint \{f(\mathbf{x}) - t\}^2 p(\mathbf{x}, t)\,\mathrm{d}\mathbf{x}\,\mathrm{d}t$$

$$\begin{aligned}\{f(\mathbf{x}) - t\}^2 &= \{f(\mathbf{x}) - \mathbb{E}[t|\mathbf{x}] + \mathbb{E}[t|\mathbf{x}] - t\}^2 \\ &= \{f(\mathbf{x}) - \mathbb{E}[t|\mathbf{x}]\}^2 + 2\{f(\mathbf{x}) - \mathbb{E}[t|\mathbf{x}]\}\{\mathbb{E}[t|\mathbf{x}] - t\} + \{\mathbb{E}[t|\mathbf{x}] - t\}^2\end{aligned}$$

$$\mathbb{E}[L] = \int \{f(\mathbf{x}) - \mathbb{E}[t|\mathbf{x}]\}^2 p(\mathbf{x})\,\mathrm{d}\mathbf{x} + \int \mathrm{var}\,[t|\mathbf{x}]\,p(\mathbf{x})\,\mathrm{d}\mathbf{x}$$

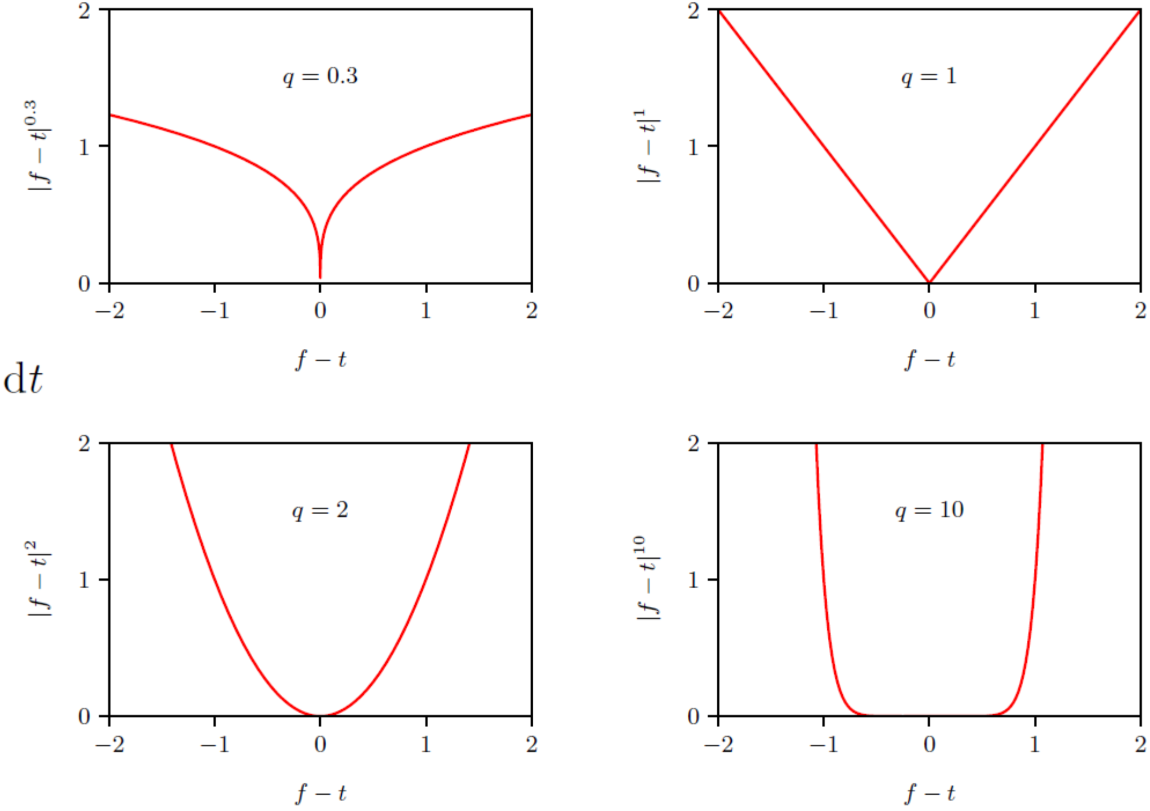
$\mathbb{E}[L]$ は $f(\mathbf{x}) = \mathbb{E}[t|\mathbf{x}]$ のとき最小.

$\mathbb{E}[L]$ is minimized when $f(\mathbf{x}) = \mathbb{E}[t|\mathbf{x}]$.

決定理論 Decision theory

ミンコフスキー損失
Minkowski loss

$$\mathbb{E}[L_q] = \iint |f(\mathbf{x}) - t|^q p(\mathbf{x}, t)\,\mathrm{d}\mathbf{x}\,\mathrm{d}t$$



バイアス–バリアンスのトレードオフ The bias–variance trade-off

期待二乗損失関数は
Recall the expected squared loss function

$$\mathbb{E}[L] = \int \{f(\mathbf{x}) - h(\mathbf{x})\}^2 p(\mathbf{x})\,\mathrm{d}\mathbf{x} + \underline{\iint \{h(\mathbf{x}) - t\}^2 p(\mathbf{x}, t)\,\mathrm{d}\mathbf{x}\,\mathrm{d}t}$$

ただし
where $h(\mathbf{x}) = \mathbb{E}[t|\mathbf{x}] = \int t p(t|\mathbf{x})\,\mathrm{d}t$

$\mathbb{E}[L]$ の第二項は t に内在するノイズに相当.
The second term of $\mathbb{E}[L]$ corresponds to the noise inherent in t .

第一項は？
What about the first term?

バイアス–バリアンスのトレードオフ The bias–variance trade-off

$$\{f(\mathbf{x};\mathcal{D}) - h(\mathbf{x})\}^2$$

あるデータ集合 $\mathcal{D}$ に対して
for a particular data set $\mathcal{D}$

$$\longrightarrow \quad \{f(\mathbf{x};\mathcal{D}) - \mathbb{E}_{\mathcal{D}}[f(\mathbf{x};\mathcal{D})] + \mathbb{E}_{\mathcal{D}}[f(\mathbf{x};\mathcal{D})] - h(\mathbf{x})\}^2$$
$$= \{f(\mathbf{x};\mathcal{D}) - \mathbb{E}_{\mathcal{D}}[f(\mathbf{x};\mathcal{D})]\}^2 + \{\mathbb{E}_{\mathcal{D}}[f(\mathbf{x};\mathcal{D})] - h(\mathbf{x})\}^2$$
$$+ 2\{f(\mathbf{x};\mathcal{D}) - \mathbb{E}_{\mathcal{D}}[f(\mathbf{x};\mathcal{D})]\}\{\mathbb{E}_{\mathcal{D}}[f(\mathbf{x};\mathcal{D})] - h(\mathbf{x})\}.$$

期待二乗損失関数の第一項
The first term of the expected squared loss function

$$\mathbb{E}_{\mathcal{D}}\left[\{f(\mathbf{x};\mathcal{D}) - h(\mathbf{x})\}^2\right]$$
$$= \underbrace{\{\mathbb{E}_{\mathcal{D}}[f(\mathbf{x};\mathcal{D})] - h(\mathbf{x})\}^2}_{(\text{bias})^2} + \underbrace{\mathbb{E}_{\mathcal{D}}\left[\{f(\mathbf{x};\mathcal{D}) - \mathbb{E}_{\mathcal{D}}[f(\mathbf{x};\mathcal{D})]\}^2\right]}_{\text{variance}}$$

バイアス–バリアンスのトレードオフ The bias–variance trade-off

期待損失＝(バイアス)^2＋バリアンス＋ノイズ

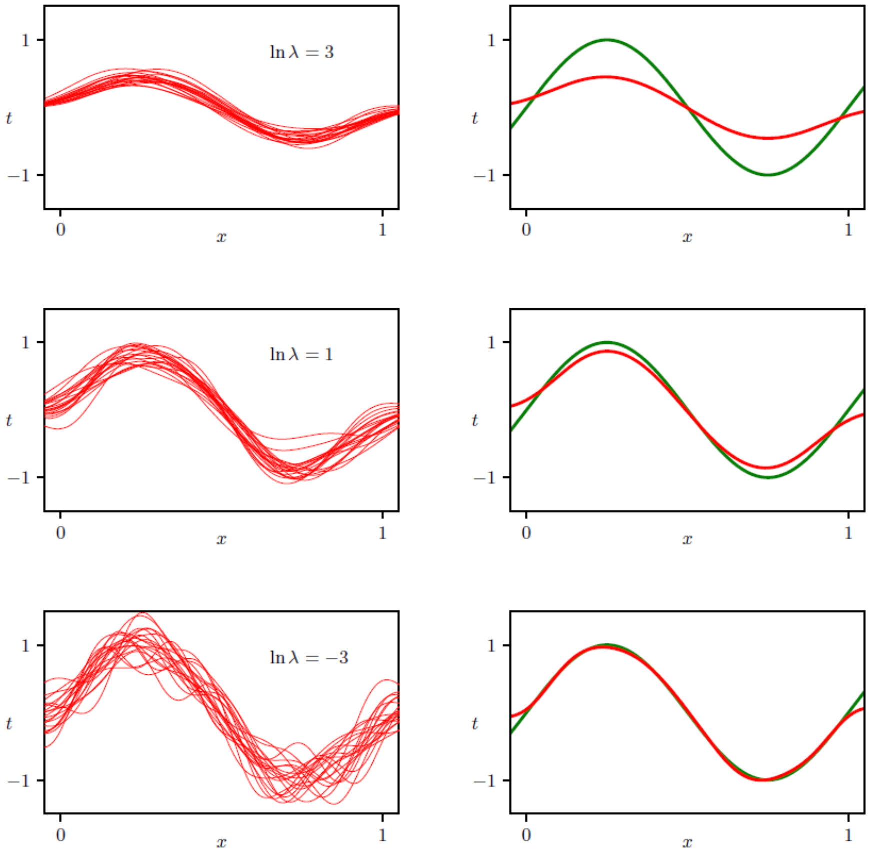
$$\text{expected loss} = (\text{bias})^2 + \text{variance} + \text{noise}$$

$$(\text{bias})^2 = \int \{\mathbb{E}_{\mathcal{D}}[f(\mathbf{x};\mathcal{D})] - h(\mathbf{x})\}^2 p(\mathbf{x})\,\mathrm{d}\mathbf{x}$$

$$\text{variance} = \int \mathbb{E}_{\mathcal{D}}\left[\{f(\mathbf{x};\mathcal{D}) - \mathbb{E}_{\mathcal{D}}[f(\mathbf{x};\mathcal{D})]\}^2\right] p(\mathbf{x})\,\mathrm{d}\mathbf{x}$$

$$\text{noise} = \iint \{h(\mathbf{x}) - t\}^2 p(\mathbf{x}, t)\,\mathrm{d}\mathbf{x}\,\mathrm{d}t$$

バイアス–バリアンスのトレードオフ The bias–variance trade-off



バイアスが大きくバリアンスが小さい
High bias and low variance

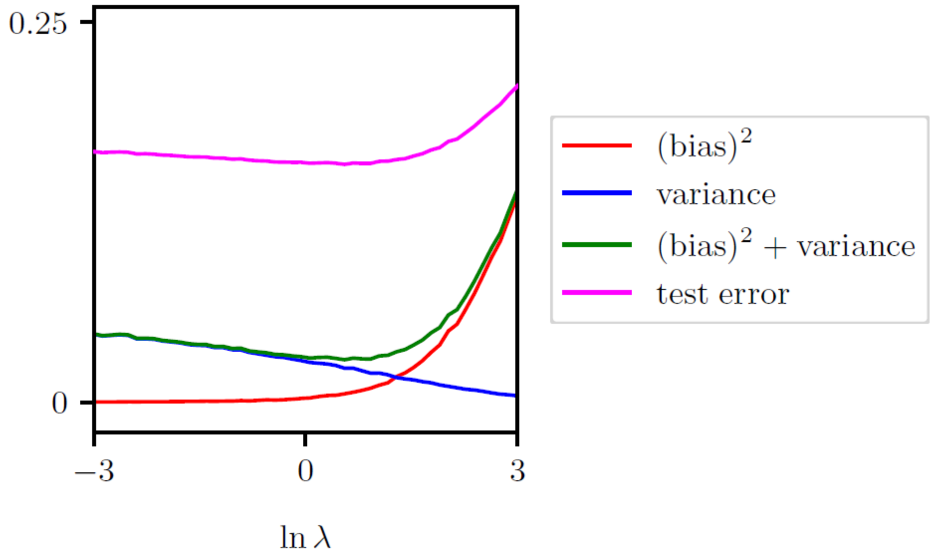
バイアスが小さくバリアンスが大きい
Low bias and high variance

バイアス–バリアンスのトレードオフ The bias–variance trade-off

$$(\text{bias})^2 = \frac{1}{N}\sum_{n=1}^{N}\left\{\overline{f}(x_n) - h(x_n)\right\}^2$$

$$\text{variance} = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{L}\sum_{l=1}^{L}\left\{f^{(l)}(x_n) - \overline{f}(x_n)\right\}^2$$

ただし
where $\overline{f}(x) = \frac{1}{L}\sum_{l=1}^{L} f^{(l)}(x)$



提出課題 II：線形回帰モデル
Assignment II: Linear Models for Regression

提出期限：**10月29日（火曜日）23:59:00** **[日本標準時]**
Submission deadline: **October 29 (Tuesday) 23:59:00** **[Japan Standard Time]**

提出課題は「一般」チャネルの「ファイル」にアップロードされます.
同チャネルに出現する通知のリンク先から解答を送信（提出）してください.
Assignments will be uploaded to "File" in the "General" channel. Send (submit) your answers via the link that will appear in the same channel.

- 全6回の課題への解答をもとに成績評価を行います.
  Your grades will be based on your answers to all six assignments.
- 解答送信後の解答再送信はできません.
  Once submitted, answers cannot be resubmitted.
- 提出期限は一切延長しません.
  The submission deadline will never be extended.