1 課題 1

1.1 問題

時間的、あるいは空間的構造を有するデータを用いて回帰を行うことを考える。このようなデータは独立性を持たず、単純にランダムサンプリングを行うことでデータが有する構造が破壊されてしまう。この時、クロスバリデーションをどのように行えばよいか考察せよ。

1.2 回答

leave-one-out クロスバリデーション (LOOCV) を行うことを考える。LOOCV は、データセットから 1 つのデータを取り出してテストデータとし、それ以外のデータを訓練データとして用いる検証である。時間的構造を持つデータであれば、テストデータとしてデータを 1 つ選択し、それよりも前のデータのみから一定数データを取り出し訓練データとして用いる。すなわち、選択した 1 つのデータ (テストデータ) を目的変数、過去のデータ (訓練データ) を説明変数として回帰モデルを構築することで、時間的構造を破壊することなく検定が行えると考える (図 1)。

また、空間的構造を持つデータにおいても、テストデータとしてデータを 1 つ選択し目的変数、その周囲の一定範囲のデータを説明変数として回帰モデルを構築することで空間的構造を破壊することなく検定が行えると考える。

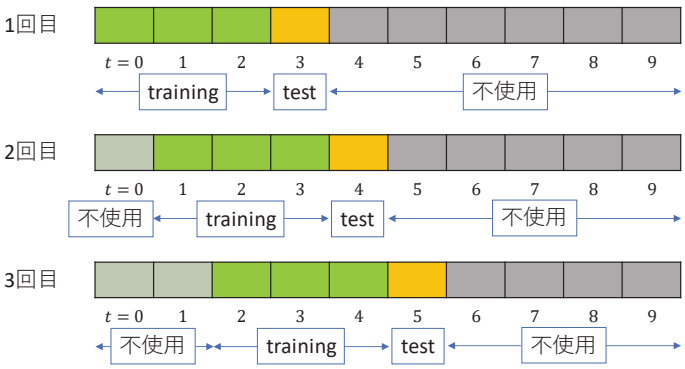

図 1 時間的構造を考慮したクロスバリデーション (説明変数数: 3)

2　課題 2

2.1　問題

式 1 のアンサンブル学習の二乗誤差のバイアス-分散-共分散分解の式を証明せよ。

$$err(H) = \bar{b}(H)^2 + \frac{1}{B}\bar{v}(H) + \Big(1 - \frac{1}{B}\Big)\bar{cv}(H) \tag{1}$$

2.2　証明

2.2.1　各変数の説明

アンサンブル学習では、B 個の仮説 h を構成する。θ_b はパラメタである。

$$h(x;\theta_b); b = 1, \ldots, B$$

今回は、同じ重みで B 個の学習機を重ねた平均で予測器 $H = \bar{h}$ を構成する。したがって、最終的な仮説 $\bar{h}$ は式 2 で構成される。

$$\bar{h}(x) = \frac{1}{B}\sum_{b=1}^{B} h(x;\theta_b) \tag{2}$$

平均化バイアス $\bar{b}(H)$、平均化分散 $\bar{v}(H)$、平均化共分散 $\bar{cv}(H)$ はそれぞれ式 3、4、5 のように定義される。

$$\bar{b}(H) = \frac{1}{B}\sum_{i=1}^{B}(E[h_i] - y) \tag{3}$$

$$\bar{v}(H) = \frac{1}{B}\sum_{i=1}^{B} E[(h_i - E[h_i])^2] \tag{4}$$

$$\bar{cv}(H) = \frac{1}{B(B-1)}\sum_{i=1}^{B}\sum_{j=1,j\neq i}^{B} E[(h_i - E[h_i])]E[(h_j - E[h_j])] \tag{5}$$

また、一般的な機械学習の平均二乗誤差は式 6 のように表される。

$$err(h) = E[(h-y)^2] = (E[h]-y)^2 + E[(h - E[h])^2] = b(h)^2 + v(h) \tag{6}$$

ただし、$b(h)$ はバイアス、$v(h)$ は分散であり、式 7, 8 のように定義される。

$$b(h) = E_{(x)}[h] - y \tag{7}$$

$$v(h) = E[(h - E_X[h])^2] \tag{8}$$

2.2.2　証明

今回は、式 6 を用いてアンサンブル学習の二乗誤差を求める。まず、$err(\cdot)$ に $H = \bar{h}$ を代入する。

$$\begin{aligned} err(H) &= err(\bar{h}) \\ &= E[(\bar{h} - y)^2] \\ &= (E[\bar{h}] - y)^2 + E[(\bar{h} - E[\bar{h}])^2] \end{aligned} \tag{9}$$

式 9 の第 1 項は、式 2、11 より以下のように変形できる。

$$(E[\bar{h}]-y)^2=\left\{\frac{1}{B}\sum_{i=1}^{B}E[h_i]-\frac{1}{B}\sum_{i=1}^{B}y\right\}^2=\left\{\frac{1}{B}\sum_{i=1}^{B}(E[h_i]-y)\right\}^2=\bar{b}(H)^2 \tag{10}$$

$$E[\bar{h}]=E\left[\frac{1}{B}\sum_{i=1}^{B}h_i\right]=\frac{1}{B}\sum_{i=1}^{B}E[h_i] \tag{11}$$

また、式 9 の第 2 項は、式 4、11 より以下のように変形できる。

$$\begin{aligned}
E[(\bar{h}-E[\bar{h}])^2] &= E\left[\left\{\frac{1}{B}\sum_{i=1}^{B}(h_i-E[h_i])\right\}^2\right] \\
&= E\left[\frac{1}{B}\sum_{i=1}^{B}(h_i-E[h_i])\cdot\frac{1}{B}\sum_{j=1}^{B}(h_j-E[h_j])\right] \\
&= \frac{1}{B^2}\sum_{i=1}^{B}\sum_{j=1}^{B}E[(h_i-E[h_i])(h_j-E[h_j])] \\
&= \frac{1}{B^2}\left\{\sum_{i=1}^{B}\sum_{j=i}^{i}E[(h_i-E[h_i])(h_j-E[h_j])]\right\} \\
&+ \frac{1}{B^2}\sum_{i=1}^{B}\sum_{j=1,j\neq i}^{B}E[(h_i-E[h_i])(h_j-E[h_j])] \\
&= \frac{1}{B}\left\{\frac{1}{B}\sum_{i=1}^{B}E[(h_i-E[h_i])^2]\right\}+\frac{1}{B^2}\sum_{i=1}^{B}\sum_{j=1,j\neq i}^{B}E[(h_i-E[h_i])(h_j-E[h_j])] \\
&= \frac{1}{B}\bar{v}(H)+\frac{1}{B^2}\sum_{i=1}^{B}\sum_{j=1,j\neq i}^{B}E[(h_i-E[h_i])(h_j-E[h_j])]
\end{aligned} \tag{12}$$

h_i はそれぞれ独立した仮説であり共分散は 0 であるため、式 5 より式 12 の第 2 項は以下のように変形できる。

$$\begin{aligned}
\frac{1}{B^2}\sum_{i=1}^{B}\sum_{j=1,j\neq i}^{B}E[(h_i-E[h_i])(h_j-E[h_j])] &= \frac{1}{B^2}\sum_{i=1}^{B}\sum_{j=1,j\neq i}^{B}E[(h_i-E[h_i])]E[(h_j-E[h_j])] \\
&= \frac{B-1}{B}\frac{1}{B(B-1)}\sum_{i=1}^{B}\sum_{j=1,j\neq i}^{B}E[(h_i-E[h_i])]E[(h_j-E[h_j])] \\
&= \left(1-\frac{1}{B}\right)\frac{1}{B(B-1)}\sum_{i=1}^{B}\sum_{j=1,j\neq i}^{B}E[(h_i-E[h_i])]E[(h_j-E[h_j])] \\
&= \left(1-\frac{1}{B}\right)\bar{cv}(H)
\end{aligned} \tag{13}$$

式 12、13 より、以下の式が得られる。

$$E[(\bar{h}-E[\bar{h}])^2]=\frac{1}{B}\bar{v}(H)+\left(1-\frac{1}{B}\right)\bar{cv}(H) \tag{14}$$

よって、式 10、14 より、式 1 が得られる。

$$\begin{aligned}
err(H) &= (E[\bar{h}]-y)^2+E[(\bar{h}-E[\bar{h}])^2] \\
&= \bar{b}(H)^2+\frac{1}{B}\bar{v}(H)+\left(1-\frac{1}{B}\right)\bar{cv}(H)
\end{aligned}$$

□