Stein's unbiased risk estimator (SURE)の解説

宮田高道

2018/07/20

1 準備

本稿を通じて，ベクトルは太字で$\mathbf{x}$のように書き，そのi番目の成分は斜体でx_iと書く．いま未知の原画像$\hat{\mathbf{x}} \in \mathbb{R}^N$に対して，（画素ごとに）平均0，分散$\sigma^2$の正規分布にしたがって発生したノイズ$\mathbf{n}$が加算された観測$\mathbf{y}$が得られているとする．すなわち

$$\mathbf{y} = \hat{\mathbf{x}} + \mathbf{n} \tag{1}$$

が成立しているとする．

ある写像$\mathbf{f} : \mathbb{R}^N \to \mathbb{R}^N$によって，観測$\mathbf{y}$から$\hat{\mathbf{x}}$の推定値$\mathbf{f}(\mathbf{y})$が得られるとき，その推定値と$\hat{\mathbf{x}}$の2乗誤差の期待値，すなわち

$$\mathbb{E}\left[||\hat{\mathbf{x}} - \mathbf{f}(\mathbf{y})||_2^2\right] \tag{2}$$

を原画像を参照することなく得ることは可能だろうか？それを実現するのがStein's unbiased risk estimator (SURE)である．

2 SUREの解説

2.1 まずは$\hat{\mathbf{x}}$を削除

いま式(1)より$\hat{\mathbf{x}} = \mathbf{y} - \mathbf{n}$である．これを式(2)に代入することにより

$$\begin{aligned}\mathbb{E}\left[||\hat{\mathbf{x}} - \mathbf{f}(\mathbf{y})||_2^2\right] &= \mathbb{E}\left[||\mathbf{y} - \mathbf{n} - \mathbf{f}(\mathbf{y})||_2^2\right] &(3)\\ &= \mathbb{E}\left[||\mathbf{y} - \mathbf{f}(\mathbf{y}) - \mathbf{n}||_2^2\right] &(4)\end{aligned}$$

が得られる．推定と観測との間の残差（すなわちノイズの推定値）$\mathbf{y} - \mathbf{f}(\mathbf{y})$をひとまとめにしたうえで2乗誤差の計算を展開すると，上式は

$$\begin{aligned}\mathbb{E}\left[||\hat{\mathbf{x}} - \mathbf{f}(\mathbf{y})||_2^2\right] &= \mathbb{E}\left[||\mathbf{y} - \mathbf{f}(\mathbf{y})||_2^2\right] + \mathbb{E}\left[||\mathbf{n}||_2^2\right] - 2\mathbb{E}\left[\mathbf{n}^\top(\mathbf{y} - \mathbf{f}(\mathbf{y}))\right] &(5)\\ &= \mathbb{E}\left[||\mathbf{y} - \mathbf{f}(\mathbf{y})||_2^2\right] + N\sigma^2 - 2\mathbb{E}\left[\mathbf{n}^\top(\mathbf{y} - \mathbf{f}(\mathbf{y}))\right] &(6)\end{aligned}$$

となる（ノイズ$\mathbf{n}$は画素ごとに独立な分散σ^2の正規分布に従っていたことを思い出そう）．この式の第1項，第2項は計算できるので，問題になるのは第3項のみである．

2.2 Steinの補題

前節の第3項，すなわち$2\mathbb{E}\left[\mathbf{n}^\top(\mathbf{y}-\mathbf{f}(\mathbf{y}))\right]$を計算するためには，以下の補題が便利である．

Theorem 2.1 Steinの補題

平均μ，分散σ^2の正規分布に従う確率変数Xと微分できる関数gについて以下が成り立つ．

$$\mathbb{E}\left[g(X)(X-\mu)\right] = \sigma^2 \mathbb{E}\left[\frac{\partial g(X)}{\partial X}\right] \tag{7}$$

この補題を適用することにより，

$$\begin{aligned}
\mathbb{E}\left[\mathbf{n}^\top(\mathbf{y}-\mathbf{f}(\mathbf{y}))\right] &= 2\mathbb{E}\left[\sum_{i=1}^{N} n_i(y_i - f(\mathbf{y})_i)\right] & (8)\\
&= 2\sigma^2\mathbb{E}\left[\sum_{i=1}^{N}\left(\frac{\partial(y_i - f(\mathbf{y})_i)}{\partial n_i}\right)\right] & (9)\\
&= 2\sigma^2\mathbb{E}\left[\sum_{i=1}^{N}\left(\frac{\partial y_i}{\partial n_i} - \frac{\partial f(\mathbf{y})_i}{\partial n_i}\right)\right] & (10)\\
&= 2\sigma^2\mathbb{E}\left[\sum_{i=1}^{N}\left(\frac{\partial y_i}{\partial n_i} - \frac{\partial f(\mathbf{y})_i}{\partial y_i}\frac{\partial y_i}{\partial n_i}\right)\right] & (11)
\end{aligned}$$

が得られる．なお，第1式と第2式の間でSteinの補題を使っており，第3式と第4式の間では微分の連鎖則を用いた．式(1)より$\frac{\partial y_i}{\partial n_i} = 1$であるから，これを代入して

$$\begin{aligned}
\mathbb{E}\left[\mathbf{n}^\top(\mathbf{y}-\mathbf{f}(\mathbf{y}))\right] &= 2\sigma^2\mathbb{E}\left[\sum_{i=1}^{N}\left(\frac{\partial y_i}{\partial n_i} - \frac{\partial f(\mathbf{y})_i}{\partial y_i}\frac{\partial y_i}{\partial n_i}\right)\right] & (12)\\
&= 2\sigma^2\mathbb{E}\left[\sum_{i=1}^{N}\left(1 - \frac{\partial f(\mathbf{y})_i}{\partial y_i}\right)\right] & (13)\\
&= 2N\sigma^2 - 2\sigma^2\mathbb{E}\left[\sum_{i=1}^{N}\left(\frac{\partial f(\mathbf{y})_i}{\partial y_i}\right)\right] & (14)
\end{aligned}$$

が得られる．

2.3 SURE

式(6)に式(14)を代入することにより，

$$\begin{aligned}
\mathbb{E}\left[||\hat{\mathbf{x}}-\mathbf{f}(\mathbf{y})||_2^2\right] &= \mathbb{E}\left[||\mathbf{y}-\mathbf{f}(\mathbf{y})||_2^2\right] + N\sigma^2 - \left(2N\sigma^2 - 2\sigma^2\mathbb{E}\left[\sum_{i=1}^{N}\left(\frac{\partial f(\mathbf{y})_i}{\partial y_i}\right)\right]\right) \\
&= \mathbb{E}\left[||\mathbf{y}-\mathbf{f}(\mathbf{y})||_2^2\right] - N\sigma^2 + 2\sigma^2\mathbb{E}\left[\sum_{i=1}^{N}\left(\frac{\partial f(\mathbf{y})_i}{\partial y_i}\right)\right] & (15)
\end{aligned}$$

が得られる．これは原画像$\hat{x}$とその推定値$\mathbf{f}(\mathbf{y})$との2乗誤差の不偏推定となることが知られており，SUREとよばれる．

なお，著者によって第3項の一部$\mathbb{E}\left[\sum_{i=1}^{N}\left(\frac{\partial f(\mathbf{y})_i}{\partial y_i}\right)\right]$を自由度 (degree of freedom, df)とよんで $\mathrm{df}_{\mathbf{f}}(\mathbf{y})$と書いたり，同じものを$\mathrm{div}(\mathbf{f})(\mathbf{y})$や$\mathrm{tr}(\partial\mathbf{f}(\mathbf{y}))$と書いたりすることがあるので注意が必要．自由度を使えばSUREは

$$\mathbb{E}\left[||\hat{\mathbf{x}}-\mathbf{f}(\mathbf{y})||_2^2\right] = \mathbb{E}\left[||\mathbf{y}-\mathbf{f}(\mathbf{y})||_2^2\right] - N\sigma^2 + 2\sigma^2\mathrm{df}_{\mathbf{f}}(\mathbf{y}) \tag{16}$$

とすっきり書ける．

3 おまけ

なお，残差$\mathbf{r}(\mathbf{y}) = \mathbf{f}(\mathbf{y}) - \mathbf{y}$に着目した場合，SUREは

$$\mathbb{E}\left[||\hat{\mathbf{x}}-\mathbf{f}(\mathbf{y})||_2^2\right] = \mathbb{E}\left[||\mathbf{r}(x)|_2^2\right] + N\sigma^2 + 2\sigma^2\mathbb{E}\left[\sum_{i=1}^{N}\left(\frac{\partial r(\mathbf{y})_i}{\partial y_i}\right)\right] \tag{17}$$

となる（導出は読者の演習問題とする）．