ベイズ最適化の指標

柴山 翔二郎

2020/06/15

1　期待改善度

期待改善度 (expected improvement, EI) は改善度の期待値を取ることで計算される．改善の確率 (probability of improvement) だけだと，本当に改善するのか否かが評価できない．

期待改善度は以下のように定義される．

$$EI(\boldsymbol{x}^*) = \begin{cases} \displaystyle\int_{-\infty}^{\infty} \max(0, f-y) N(y|\mu, \sigma^2) dy & \text{（最小化）} \\ \displaystyle\int_{-\infty}^{\infty} \max(0, y-f) N(y|\mu, \sigma^2) dy & \text{（最大化）} \end{cases} \tag{1}$$

ここで f は最適値を示す．また，y, μ, σ は $\boldsymbol{x}^*$ に対するベイズ過程回帰の結果である．

1.1　y の最小化

以下では式 (1) を計算する．それぞれの場合で $y = \mu + \sigma\epsilon$ ($\epsilon \sim N(\epsilon|0,1)$) とする．

$$\begin{aligned} EI(\boldsymbol{x}^*) &= \int_{-\infty}^{f} (f-y) N(y|\mu, \sigma^2) dy \\ &= \int_{-\infty}^{(f-\mu)/\sigma} (f-\mu-\epsilon\sigma) N(\epsilon|0,1) d\epsilon \\ &= (f-\mu) \int_{-\infty}^{(f-\mu)/\sigma} N(\epsilon|0,1) d\epsilon + \sigma \int_{-\infty}^{(f-\mu)/\sigma} (-\epsilon) N(\epsilon|0,1) d\epsilon \\ &= (f-\mu)\Phi\left(\frac{f-\mu}{\sigma}\middle| 0,1\right) + \sigma\phi\left(\frac{f-\mu}{\sigma}\middle| 0,1\right) \end{aligned} \tag{2}$$

ここで Φ は累積密度関数 (cumlative density function, CDF)，ϕ は確率密度関数 (probability density function, PDF) を指す．

1.2 y の最大化

$$
\begin{aligned}
EI(\boldsymbol{x}^*) &= \int_f^\infty (y-f)N(y,\mu,\sigma^2)dy \\
&= \int_{(f-\mu)/\sigma}^\infty (\mu+\sigma\epsilon-f)N(\epsilon|0,1)d\epsilon \\
&= (\mu-f)\int_{(f-\mu)/\sigma}^\infty N(\epsilon|0,1)d\epsilon - \sigma\int_{(f-\mu)/\sigma}^\infty (-\epsilon)N(\epsilon|0,1)d\epsilon \\
&= (\mu-f)\int_{(f-\mu)/\sigma}^\infty N(\epsilon|0,1)d\epsilon - \sigma\left[\frac{1}{Z}\exp\left(\frac{\epsilon^2}{2}\right)\right]_{(f-\mu)/\sigma}^\infty \\
&= (\mu-f)\left(1-\Phi\left(\frac{f-\mu}{\sigma}\middle|0,1\right)\right) + \sigma\phi\left(\frac{f-\mu}{\sigma}\middle|0,1\right) \\
&= (\mu-f)\Phi\left(\frac{\mu-f}{\sigma}\middle|0,1\right) + \sigma\phi\left(\frac{\mu-f}{\sigma}\middle|0,1\right)
\end{aligned}
\tag{3}
$$

ここでも Φ で CDF を，ϕ で PDF を指す．また，式の途中で $Z=1/\sqrt{2\pi\sigma^2}$ としてガウス分布の正規化定数とした．CDF は $-\infty$ から確率密度関数を積分した値なので，最大化の EI の式を踏まえると，1 から CDF を引いた値を第一項に持ってくる必然性が理解できる．また，式 (2) と (3) を比べると第二項は同じであることが分かる．

一番最後の行ではガウス分布の対称性を利用した．つまり，$f-\mu$ の符号を変換するだけで画一的に EI を書き下すことができる．

$$
\begin{aligned}
\Phi\left(\frac{f-\mu}{\sigma}\middle|0,1\right) &= \Phi\left(f|\mu,\sigma\right) \\
\phi\left(\frac{f-\mu}{\sigma}\middle|0,1\right) &= \phi\left(f|\mu,\sigma^2\right)
\end{aligned}
\tag{4}
$$

が成り立つはずである．（このことを利用すると実装は非常に楽になるはず．）

1.3 式変形による簡便な表記

@t_inoue 氏の指摘をもらい以下のような表記の簡略化が可能である．まず確率密度関数については以下のように直される．実装時に注意が必要となる．

$$
\begin{aligned}
\phi\left(\tfrac{f-\mu}{\sigma}\middle|0,1\right) &= \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{f-\mu}{\sigma}\right)^2\right) \\
&= \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2\sigma^2}\left(f-\mu\right)^2\right) \\
&= \frac{\sigma}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}\left(f-\mu\right)^2\right) \\
&= \sigma\phi(f|\mu,\sigma^2)
\end{aligned}
\tag{5}
$$

次に $\Phi\left(\frac{f-\mu}{\sigma}\middle|0,1\right)$，$\phi\left(\frac{f-\mu}{\sigma}\middle|0,1\right)$ の対称性を考えると次の式が成り立つ．

$$
\begin{aligned}
\phi\left(\frac{f-\mu}{\sigma}\middle|0,1\right) &= \phi\left(\frac{f-\mu}{\sigma}\middle|0,1\right) \\
\Phi\left(\frac{f-\mu}{\sigma}\middle|0,1\right) &= 1-\Phi\left(\frac{\mu-f}{\sigma}\middle|0,1\right)
\end{aligned}
\tag{6}
$$

式 (2)，(3) を式 (6) を使い式変形して簡潔に表現すると次のようになる．

$$
\begin{aligned}
EI(\boldsymbol{x}^*) &= Z\Phi\left(\frac{Z}{\sigma}\middle|\, 0,1\right) + \sigma\phi\left(\frac{Z}{\sigma}\middle|\, 0,1\right) \\
\text{ただし} \\
Z &= \begin{cases} f-\mu & \text{最小化} \\ \mu - f & \text{最大化} \end{cases}
\end{aligned}
\tag{7}
$$

2　まとめ

期待改善度を導出した．変数変換により見通しを良くして，累積密度関数と確率密度関数を含んだ式を示した．

ベイズ最適化の記事ではほとんどの場合目的変数最小化を目指していて EI についても最小化指標しか示されていない．今回は最大化の際の式も導出したので，以降は上式に従い実装をすると良い．