1 シグモイドカーネルの半正定値性

シグモイドカーネルによって得られるグラム行列は半正定値性を満たさないという言説は多く見かけるが, その証明を目にすることは少ない. ここではシグモイドカーネルが半正定値性を満たす条件を考えよう.

ちなみに筆者ははじめ, テイラー展開や多項式カーネルとの関係を駆使して証明を試みたがどうにも上手くいかなかった. ここはおとなしく先人の知恵を借りて, [1] の翻訳と行間の補充に徹することにする.

シグモイドカーネルは次式で与えられる.

$$K(x_i, x_j) = tanh(ax_i^t x_j + b) \tag{1}$$

1.1 $b < 0$

シグモイド関数の定義より, 次の等式が成り立つ.

$$\frac{1}{1+e^{-2x}} = \frac{1+tanh(x)}{2} \tag{2}$$

さて, 次のような式を考えよう.

$$\lim_{x\to-\infty} \frac{1+tanh(x+\delta)}{1+tanh(x)} \tag{3}$$

式 (2)(3) より

$$\begin{aligned}
\lim_{x\to-\infty} \frac{1+tanh(x+\delta)}{1+tanh(x)} &= \lim_{x\to-\infty} \frac{1+e^{-2x}}{1+e^{-2(x+\delta)}} && (4)\\
&= \lim_{x\to-\infty} \frac{1+e^{-2x}}{1+e^{-2x}e^{-2\delta}} && (5)\\
&= e^{2\delta} && (6)
\end{aligned}$$

式 (4) の左辺の x に b を, δ に $ax_i^t x_j$ を代入して

$$\lim_{b\to-\infty} \frac{1+tanh(b+ax_i^t x_j)}{1+tanh(b)} = e^{2ax_i^t x_j} \tag{7}$$

さて, 余弦定理は次式で与えられる.

$$||x_i - x_j||_2^2 = ||x_i||_2^2 + ||x_j||_2^2 - 2x_i^t x_j \tag{8}$$

これを用いて, 式 (7) を次のように変形する.

$$e^{2ax_i^t x_j} = e^{a||x_i||_2^2} e^{-a||x_i-x_j||_2^2} e^{a||x_j||_2^2} \tag{9}$$

ここで, $e^{-a||x_i-x_j||_2^2}$ は RBF カーネルに一致する. $e^{a||x_i||_2^2}$ と $e^{a||x_j||_2^2}$ は常に正だから, 式 (9) は RBF カーネルの正倍である. よって a が正で b が十分に小さい時, シグモイドカーネルの与えるグラム行列は正定値である.

逆に, $a > 0$ が RBF カーネルが正定値性を満たす条件であるから, $a \leq 0$ の時, それを正倍して得られるシグモイドカーネルは正定値性を満たさない.

1.2 $b > 0$

Merser の定理より, カーネル関数によって与えられるグラム行列が正定値性を満たすなら次式が成り立つ.

$$K(x_i, x_i) + K(x_j, x_j) - 2K(x_i, x_j) > 0 \tag{10}$$

シグモイドカーネルにおける不等式は具体的に次式で与えられる.

$$tanh(ax_i^t x_i + b) + tanh(ax_j^t x_j + b) - 2tanh(ax_i^t x_j + b) > 0 \tag{11}$$

さて, $max(ax_i^t x_i + b) < 0$ を満たす時, 次式が成り立つ.

$$\frac{tanh(ax_i^t x_i + b) + tanh(ax_j^t x_j + b)}{2} \quad \geq \quad tanh(\frac{ax_i^t x_i + b + ax_j^t x_j + b}{2}) \tag{12}$$

この時, 適当な $x_i, x_j < 0$ に対して

$$\frac{tanh(x_i) + tanh(x_j)}{2} \quad \geq \quad tanh(\frac{x_i + x_j}{2}) \tag{13}$$

これは自明ではないが, 例えば x_i に -100, x_j に -0.1 を代入してみるとわかり易い. 同様に 100 と 0.1 を代入してみると不等式が成り立たないことがわかる.

さて, 式 (12) は更に

$$\begin{aligned} tanh(\frac{ax_i^t x_i + b + ax_j^t x_j + b}{2}) &= tanh(a\frac{x_i^t x_i + x_j^t x_j}{2} + b) && (14) \\ &> tanh(ax_i^t x_j + b) && (15) \end{aligned}$$

式 (15) は $i \neq j$ であるが, その時コサイン類似度は 1 未満であることから自明である. よって式 (11) から (15) により (10) が満たされた.

しかし, 任意の x_i に対して, $max(ax_i^t x_i + b) < 0$ は $a > 0$ の時 $b < 0$ でなければ満たされない.

$a < 0$ の時は (14) と (15) の不等号が逆になり, (10) を満たせない.

結局, シグモイドカーネルでは $a > 0, b << 0$ で正定値性を満たすことがわかる. しかし b の小ささの度合いについては x_i のスケーリングと相まってはじめから定義できない. このあたりが半正定値性を満たさないといわれる所以だろう.

2018.2.9 筑波大学 佐藤怜

参考文献

[1] Hsuan-Tien Lin and Chih-Jen Lin, A Study on Sigmoid Kernels for SVM and the Training of non-PSD Kernels by SMO-type Methods, `http://home.caltech.edu/~htlin/publication/doc/tanh.pdf`