Proceeding Series of the Brazilian Society of Applied and Computational Mathematics, Vol. 5, N. 1, 2017.

Trabalho apresentado no CNMAC, Gramado - RS, 2016.

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Um novo algoritmo aplicado em Ridge Regression

Tatiane Cazarin da Silva[1]
Departamento de Matemática, UTFPR, Campo Mourão, PR
Programa de Pós-Graduação em Métodos Numéricos em Engenharia, UFPR, Curitiba, PR
Ademir Alves Ribeiro[2]
Departamento de Matemática, UFPR, Curitiba, PR
Gislaine Aparecida Periçaro[3]
Colegiado de Matemática, UNESPAR, Campo Mourão, PR

1 Introdução

Neste trabalho, estabelecemos uma formulação geral primal-dual de ponto fixo aplicada ao problema *Ridge Regression* [1,2]. Descrevemos a dualidade convexa para essa classe de métodos e propomos um algoritmo para minimizar o gap de dualidade, nomeado método SRP, estruturado a seguir.

2 O método SRP

Considere matrizes $A_1, \dots, A_n \in \mathbb{R}^{d\times m}$, vetores $y_1, \dots, y_n \in \mathbb{R}^m$ e um parâmetro de regularização $\lambda > 0$. Denotamos $A = (A_1 \ \cdots \ A_n) \in \mathbb{R}^{d\times N}$ e $y = (y_1^T \ \cdots \ y_n^T)^T \in \mathbb{R}^N$, onde $N = nm$. Neste trabalho consideramos o problema de minimização regularizada

$$\min_{w\in\mathbb{R}^d} P(w) = \frac{1}{2n}\|A^T w - y\|^2 + \frac{\lambda}{2}\|w\|^2 \tag{1}$$

e associamos o problema dual de Fenchel

$$\max_{\alpha\in\mathbb{R}^N} D(\alpha) = -\frac{1}{2\lambda n^2}\|A\alpha\|^2 - \frac{1}{2n}\|\alpha\|^2 + \frac{1}{n}\alpha^T y. \tag{2}$$

Definindo $x = \begin{pmatrix} w \\ \alpha \end{pmatrix} \in \mathbb{R}^{d+N}$, nosso problema primal-dual consiste na minimização do gap de dualidade, dado por

$$\min_{x\in\mathbb{R}^{d+N}} f(x) = P(w) - D(\alpha). \tag{3}$$

[1]tatianecazarin@utfpr.edu.br
[2]ademir.ribeiro@ufpr.br
[3]gislaine.pericaro@unespar.edu.br

Usando conceitos de dualidade convexa, a condição de otimalidade, $\nabla f(x) = 0$, com garantia de existência de solução única, recai no sistema a seguir

$$x = G(\theta)x + \theta b, \tag{4}$$

onde $G(\theta) = \begin{pmatrix} (1-\theta)\, I - \dfrac{\theta}{\lambda n} AA^T & 0 \\ 0 & \left(1 - \dfrac{\theta}{\lambda n}\right) I - \dfrac{\theta}{(\lambda n)^2} A^T A \end{pmatrix}$, $x = (w \quad \alpha)^T$ e $b = \dfrac{1}{\lambda n}\,(Ay \quad y)^T$.

Note que o formato de ponto fixo dado em (4) nos sugere, iterativamente, a sequência

$$x^{k+1} = G(\theta)x^k + \theta b.$$

Utilizando características da análise de convergência do método de ponto fixo, que baseia-se em propriedades do raio espectral da matriz $G(\theta)$, provamos o seguinte resultado.

Teorema 1. *Seja $x^0 \in \mathbb{R}^{d+N}$ um ponto inicial arbitrário e considere a sequência $(x^k)_{k\in\mathbb{N}}$ gerada pelo algoritmo com $\lambda n \geq 1$ e $\theta \in \left(0, \dfrac{2\lambda n}{\lambda n + \sigma_1^2}\right)$, onde σ_1 é o maior valor singular da matriz A. Então, a sequência (x^k) converge linearmente para a solução do problema (3), com taxa de convergência dada por*

$$\rho(\theta) = \max\left\{\left|1 - \theta - \frac{\theta\sigma_1^2}{\lambda n}\right|, \left|1 - \frac{\theta}{\lambda n}\right|\right\}.$$

Além disso, se escolhermos $\theta^* = \dfrac{2\lambda n}{\lambda n + \sigma_1^2 + 1}$, então a taxa de convergência é otima e dada por

$$\rho^* = \frac{\sigma_1^2 + \lambda n - 1}{\sigma_1^2 + \lambda n + 1}.$$

Com essa análise, estabelecemos um novo algoritmo primal-dual de ponto fixo para resolver o problema (3), onde o primal e o dual são considerados simultaneamente. A convergência teórica foi verificada e a eficiência comprovada numericamente.

Referências

[1] C. E. Ginestet, Regularization: Ridge Regression and Lasso, Notas de aula, Department of Mathematics and Statistics, Boston University, 2014, Disponível em: <http://math.bu.edu/people/cgineste/classes/ma575/p/w14_1.pdf>.

[2] A. E. Hoerl, R. W. Kennard, Ridge regression: biased estimation for nonorthogonal problems, *Technometrics*, 12 (1):55–67, 1970.