Nonnegative Least Squares

Geometry and its KKT conditions

Andersen Ang

Mathématique et recherche opérationnelle
UMONS, Belgium

manshun.ang@umons.ac.be   Homepage: angms.science

First draft : December 12, 2020
Last update : December 14, 2020

Nonnegative Least Squares

- Given $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$, find $\mathbf{x} \in \mathbb{R}^n$

$$(\mathcal{P}) \;\; : \;\; \min_{\mathbf{x} \geq 0} f(\mathbf{x}) = \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2.$$

- We will use $(\mathcal{P})$ as an example to study the KKT conditions.
- We will see that, the gradient of f is the Lagrangian multiplier.

Cone

- Given a set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ in $\mathbb{R}^m$, the cone generated by these vectors, denoted as cone$(\mathbf{v}_1, \ldots, \mathbf{v}_n)$, is defined as

$$\text{cone}(\mathbf{v}_1, \ldots, \mathbf{v}_n) = \left\{ \mathbf{x} \in \mathbb{R}^m \;\middle|\; \begin{array}{l} \mathbf{x} = \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_n \mathbf{v}_n \\ \alpha_1 \geq 0, \alpha_2 \geq 0, \ldots, \alpha_n \geq 0 \end{array} \right\}.$$

- Compact notation: let $\mathbf{V} \in \mathbb{R}^{m \times n}$ as $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$; and let $\boldsymbol{\alpha} \in \mathbb{R}^n$ be the vector representation of $\alpha_1, \ldots, \alpha_n$.

$$\text{cone}(\mathbf{V}) = \left\{ \mathbf{x} \in \mathbb{R}^m \;\middle|\; \mathbf{x} = \mathbf{V}\boldsymbol{\alpha}, \boldsymbol{\alpha} \geq 0 \right\}.$$

- The geometric meaning of cone = the collection of vectors that are nonnegative sum of the vectors $\mathbf{v}_i$.

- Other name of nonnegative sum: conical combination, conical sum.

NNLS = projection onto cone

- Given $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$, find $\mathbf{x} \in \mathbb{R}^n$

$$(\mathcal{P}) \;\;:\;\; \min_{\mathbf{x} \geq 0} f(\mathbf{x}) = \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2.$$

- Using cone geometry, NNLS is equivalent to the problem of projection onto a cone

$$(\mathcal{P}') \;\;:\;\; \min_{\mathbf{y}} f(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{b}\|_2^2 \quad \text{s.t.} \quad \mathbf{y} \in \text{cone}(\mathbf{A}),$$

  that is, we ask to find a vector $\mathbf{y}$ inside the cone generated by $\mathbf{A}$ that is closest to the given vector $\mathbf{b}$.

- Problem $\mathcal{P}'$ always has a solution.
  - cone$(\mathbf{A})$ is non-empty
  - the function $\|\mathbf{y} - \mathbf{b}\|_2^2$ is strongly-convex and thereby coercive $\implies$ solution exists and unique.

Karush-Kuhn-Tucker conditions, a review

- For a constrained problem

$$\min_{\mathbf{x}} f(\mathbf{x}) \quad \text{s.t.} \quad g(\mathbf{x}) \leq 0,$$

  where $g : \mathbb{R}^n \to \mathbb{R}^n$ and the $\leq$ is taken elementwise.

- The associated Lagrangian is

$$L(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \langle \boldsymbol{\lambda}, g(\mathbf{x}) \rangle$$

  where $\boldsymbol{\lambda} \in \mathbb{R}^n$ is the Lagrangian multiplier.

- The KKT conditions of such problem

$$\begin{aligned} \nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}) &= 0 && \text{stationarity} \\ g(\mathbf{x}) &\leq 0 && \text{Primal feasibility} \\ \boldsymbol{\lambda} &\geq 0 && \text{Dual feasibilitiy} \\ \lambda_i [g(\mathbf{x})]_i &= 0 && \text{Complementary slackness} \end{aligned}$$

KKT conditions on NNLS

$$\min_{\mathbf{x}\geq 0} f(\mathbf{x}) = \frac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2^2.$$

- The constraint $\mathbf{x} \geq 0$ can be written as $g(\mathbf{x}) \leq 0$ with $g(\mathbf{x}) = -\mathbf{I}\mathbf{x}$.
- Let $\boldsymbol{\lambda} \in \mathbb{R}^n$ be the Lagrangian multiplier, the associated Lagrangian is

$$L(\mathbf{x},\boldsymbol{\lambda}) = f(\mathbf{x}) + \langle \boldsymbol{\lambda}, g(\mathbf{x})\rangle = \frac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2^2 + \langle \boldsymbol{\lambda}, -\mathbf{I}\mathbf{x}\rangle.$$

  We have $\nabla_{\mathbf{x}} L = \mathbf{A}^\top\mathbf{A}\mathbf{x} - \mathbf{A}^\top\mathbf{b} - \boldsymbol{\lambda}$. By the KKT stationarity condition $\nabla_{\mathbf{x}} L = 0$, we have $\boldsymbol{\lambda} = \mathbf{A}^\top\mathbf{A}\mathbf{x} - \mathbf{A}^\top\mathbf{b} = \nabla_{\mathbf{x}} f(\mathbf{x})$, that is, the Lagrangian multiplier of NNLS is the gradient of f.

- The KKT conditions for NNLS now reduce to only 3 lines

$$\begin{aligned} \mathbf{x} &\geq 0 \\ \mathbf{A}^\top\mathbf{A}\mathbf{x} - \mathbf{A}^\top\mathbf{b} &\geq 0 \\ x_i[\mathbf{A}^\top\mathbf{A}\mathbf{x} - \mathbf{A}^\top\mathbf{b}]_i &= 0 \end{aligned}$$

The idea of active-set method

$$\min_{\mathbf{x}\geq 0} \frac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2^2, \quad \text{KKT} \begin{cases} \mathbf{x} \geq 0 \\ \mathbf{A}^\top\mathbf{A}\mathbf{x} - \mathbf{A}^\top\mathbf{b} \geq 0 \\ x_i[\mathbf{A}^\top\mathbf{A}\mathbf{x} - \mathbf{A}^\top\mathbf{b}]_i = 0 \end{cases}$$

- Suppose we know the support $\mathcal{S}$ of $\mathbf{x}$: i.e. the index set of $\mathbf{x}$ such that x_i is non-zero. Mathematically,

$$\mathcal{S} = \Big\{ \; i \in \{1,2,\ldots,n\}, \; | \; x_i \neq 0 \Big\}.$$

  Then by the complementary slackness condition, we know $[\mathbf{A}^\top\mathbf{A}\mathbf{x} - \mathbf{A}^\top\mathbf{b}]_{\mathcal{S}} = 0$. That is, we can then consider solving the linear system

$$\mathbf{A}_{\mathcal{S}}\mathbf{x}_{\mathcal{S}} = \mathbf{b}_{\mathcal{S}}$$

  where $\mathbf{A}$ are submatrix of $\mathbf{A}$ with row index in $\mathcal{S}$.

- The set $\mathcal{S}$ such that $\mathbf{A}_{\mathcal{S}}\mathbf{x}_{\mathcal{S}} = \mathbf{b}_{\mathcal{S}}$ is called Active-set. The idea of Active-set method is based on finding the set $\mathcal{S}$ to solve NNLS.

KKT conditions on NMF

- NMF is the problem

$$\min_{\mathbf{W}\geq 0, \mathbf{H}\geq 0} f(\mathbf{W}, \mathbf{H}) = \frac{1}{2}\|\mathbf{M} - \mathbf{W}\mathbf{H}\|_2^2.$$

- Following the same line of arguments as in NNLS, the KKT conditions for NMF are

$$\begin{array}{ll} \mathbf{W} \geq 0, & \mathbf{H} \geq 0, \\ \mathbf{W}\mathbf{H}\mathbf{H}^\top - \mathbf{M}\mathbf{H}^\top \geq 0, & \mathbf{W}^\top\mathbf{W}\mathbf{H} - \mathbf{W}^\top\mathbf{M} \geq 0, \\ W_{ij}[(\mathbf{W}\mathbf{H} - \mathbf{M})\mathbf{H}^\top]_{ij} = 0, & H_{ij}[\mathbf{W}^\top(\mathbf{W}\mathbf{H} - \mathbf{M})]_{ij} = 0. \end{array}$$

- Some authors will write down the Lagrangian multipliers $\boldsymbol{\lambda}_{\mathbf{W}}$ and $\boldsymbol{\lambda}_{\mathbf{H}}$ explicitly, which is in fact redundant because we already know the gradient $\nabla_{\mathbf{W}} f$ and $\nabla_{\mathbf{H}} f$ are the Lagrangian multipliers.

Last page - summary

- For the NNLS problem

$$\min_{\mathbf{x}\geq 0} f(\mathbf{x}) = \frac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2^2,$$

  is equivalent to project $\mathbf{b}$ onto Cone$(\mathbf{A})$.

- The KKT conditions for NNLS has only 3 lines

$$\begin{aligned} \mathbf{x} &\geq 0 \\ \mathbf{A}^\top\mathbf{A}\mathbf{x}-\mathbf{A}^\top\mathbf{b} &\geq 0 \\ x_i[\mathbf{A}^\top\mathbf{A}\mathbf{x}-\mathbf{A}^\top\mathbf{b}]_i &= 0 \end{aligned}$$

  where $\nabla_{\mathbf{x}} f(\mathbf{x})$ is the Lagrangian multiplier.

- The complementary slackness condition of NNLS is the basis of the Active-Set method for solving NNLS.

End of document