Optimisation quadratique

25 février 2021

Introduction

Définition

- un problème d'optimisation quadratique générique s'écrit

$$\begin{array}{ll} \min_{\mathbf{x}\in\mathbb{R}^d} & \frac{1}{2}\mathbf{x}^\top\mathbf{Q}\mathbf{x} + \mathbf{c}^\top\mathbf{x} \\ \text{sc} & \mathbf{A}\mathbf{x} = \mathbf{a} \\ & \mathbf{B}\mathbf{x} \leq \mathbf{b} \\ & \mathbf{x} \geq 0 \end{array}$$

  où $\mathbf{Q} \in \mathbb{R}^{d\times d}$ est une matrice symétrique et $\mathbf{A} \in \mathbb{R}^{m_1\times d}$, $\mathbf{B} \in \mathbb{R}^{m_2\times d}$. Dans notre cas, on suppose que $\mathbf{Q}$ est définie positive

- parfois on note
  - $\mathbf{A}\mathbf{x} = \mathbf{a}$ par $\mathbf{a}_i^\top\mathbf{x} = a_i$ avec $i = 1, \cdots, m_1$
  - $\mathbf{B}\mathbf{x} = \mathbf{b}$ par $\mathbf{b}_j^\top\mathbf{x} = b_j$ avec $j = 1, \cdots, m_2$

Application d'un problème QP

Régression quadratique parcimonieux

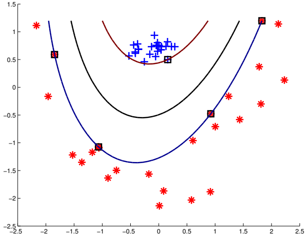
$$\begin{array}{ll} \min_{\mathbf{w},t_i} & \frac{1}{2}\|\mathbf{y}-\mathbf{X}\mathbf{w}\|_2^2 + \lambda \sum_i t_i \\ \text{sc} & -t_i \leq w_i \leq t_i \end{array}$$

Problème dual SVM

Les variables duales optimales s'obtiennent

$$\begin{array}{ll} \min_{\alpha} & \frac{1}{2}\alpha^\top \mathbf{Q}\alpha - \alpha^\top \mathbf{1} \\ \text{sc} & \alpha^\top \mathbf{y} = 0 \\ & 0 \leq \alpha_i \leq C \end{array}$$

où $Q_{i,j} = y_i y_j k(\mathbf{x}_i, \mathbf{x}_j)$



Quelques classes de QP

QP non-contraint

$$\min_{\mathbf{x}\in\mathbb{R}^d} \quad \frac{1}{2}\mathbf{x}^\top \mathbf{Q}\mathbf{x} + \mathbf{c}^\top \mathbf{x}$$

QP avec contraintes de boites

$$\begin{array}{ll} \min_{\mathbf{x}\in\mathbb{R}^d} & \frac{1}{2}\mathbf{x}^\top \mathbf{Q}\mathbf{x} + \mathbf{c}^\top \mathbf{x} \\ & l_i \leq \mathbf{x}_i \leq u_i \quad \forall i \end{array}$$

QP avec contraintes d'égalité

$$\begin{array}{ll} \min_{\mathbf{x}\in\mathbb{R}^d} & \frac{1}{2}\mathbf{x}^\top \mathbf{Q}\mathbf{x} + \mathbf{c}^\top \mathbf{x} \\ \text{sc} & \mathbf{A}\mathbf{x} = \mathbf{a} \end{array}$$

QP avec contraintes d'inégalité

$$\begin{array}{ll} \min_{\mathbf{x}\in\mathbb{R}^d} & \frac{1}{2}\mathbf{x}^\top \mathbf{Q}\mathbf{x} + \mathbf{c}^\top \mathbf{x} \\ \text{sc} & \mathbf{A}\mathbf{x} = \mathbf{a} \\ & \mathbf{B}\mathbf{x} \leq \mathbf{b} \end{array}$$

QP non-contraint

Le problème

$$\min_{\mathbf{x}\in\mathbb{R}^d} \quad \frac{1}{2}\mathbf{x}^\top \mathbf{Q}\mathbf{x} + \mathbf{c}^\top \mathbf{x}$$

Résolution

- Descente de gradient ou Méthode de Newton
- algorithme de gradient conjugué
  - algorithme trouvant un $\mathbf{x}^\star$ tel que $\mathbf{Q}\mathbf{x}^\star = -\mathbf{c}$
  - `https://www.cs.cmu.edu/~quake-papers/painless-conjugate-gradient.pdf`

QP avec contraintes de boites

Le problème

$$\begin{array}{ll} \min_{\mathbf{x}\in\mathbb{R}^d} & \frac{1}{2}\mathbf{x}^\top\mathbf{Q}\mathbf{x} + \mathbf{c}^\top\mathbf{x} \\ & l_i \leq \mathbf{x}_i \leq u_i \quad \forall i \end{array}$$

Cas d'usage

- Débruitage en image : $l_i = 0$ et $u_i = 255$.
- SVM sans biais

Résolution

- Algorithme de gradient projeté
- Projection sur une contrainte de boite

$$\mathbf{x}^\star = \arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y}-\mathbf{x}\|_2^2 \quad \text{sc} \quad l_i \leq \mathbf{x}_i \leq u_i \quad \forall i$$

$x_i^\star = y_i$ si $l_i \leq y_i \leq u_i$, $x_i^\star = l_i$ si $y_i \leq l_i$ et $x_i^\star = u_i$ si $y_i \geq u_i$

- Il existe des variantes de gradient projeté très efficace.

QP avec contraintes

Cas général

$$\begin{array}{ll} \min_{\mathbf{x}\in\mathbb{R}^d} & \frac{1}{2}\mathbf{x}^\top\mathbf{Q}\mathbf{x}+\mathbf{c}^\top\mathbf{x} \\ \text{sc} & \mathbf{A}\mathbf{x}=\mathbf{a} \\ & \mathbf{x}_i \geq 0 \quad \forall i \end{array}$$

où $\mathbf{A} \in \mathbb{R}^{m\times d}$

Transformation des cas spécifiques

- les contraintes d'inégalités peuvent etre prise en compte dans cette forme car $\mathbf{Bx} \leq \mathbf{b}$ equivaut à $\mathbf{Bx}+\mathbf{s}=\mathbf{b}$. On a donc un problème augmenté où les variables deviennent

$$\tilde{\mathbf{x}}=\begin{bmatrix}\mathbf{x}\\ \mathbf{s}\end{bmatrix} \quad \tilde{\mathbf{Q}}=\begin{bmatrix}\mathbf{Q} & 0\\ 0 & 0\end{bmatrix} \quad \tilde{\mathbf{c}}=\begin{bmatrix}\mathbf{c}\\ 0\end{bmatrix} \quad \tilde{\mathbf{A}}=\begin{bmatrix}\mathbf{A} & 0\\ \mathbf{B} & \mathbf{I}\end{bmatrix} \quad \tilde{\mathbf{a}}=\begin{bmatrix}\mathbf{a}\\ \mathbf{b}\end{bmatrix}$$

QP : algorithme de points intérieurs

Principe

- Utilisation d'une barrière $\log$ pour les contraintes de positivité.
- Le problème associé devient $(BQP)_\mu$

$$
\begin{array}{ll}
\min_{\mathbf{x}\in\mathbb{R}^d_+} & \frac{1}{2}\mathbf{x}^\top\mathbf{Q}\mathbf{x} + \mathbf{c}^\top\mathbf{x} - \mu\sum_i \log(x_i) \\
\text{sc} & \mathbf{A}\mathbf{x} = \mathbf{b}
\end{array}
$$

- $\mu\log(x_i)$ agit comme une pénalité pour les variables s'approchant de 0
- Lorsque μ est grand, la solution est à l'intérieur du domaine d'admissibilité défini par l'orthant positif et l'hyperplan $\mathbf{A}\mathbf{x} = \mathbf{b}$.

Lagrangien associé

$$
\mathcal{L}_\mu(\mathbf{x}, \lambda) = \frac{1}{2}\mathbf{x}^\top\mathbf{Q}\mathbf{x} + \mathbf{c}^\top\mathbf{x} - \mu\sum_i \log(x_i) - \lambda^\top(\mathbf{A}\mathbf{x} - \mathbf{b})
$$

QP Points intérieurs

Condition KKT associé à $(BQP)_\mu$

$$
\begin{aligned}
\frac{\partial \mathcal{L}_\mu}{\partial \mathbf{x}} &= 0 \Rightarrow \mathbf{Q}\mathbf{x} + \mathbf{c} - \mu \mathbf{X}^{-1}\mathbf{e} - \mathbf{A}^\top \lambda = 0 \\
\frac{\partial \mathcal{L}_\mu}{\partial \lambda} &= 0 \Rightarrow \mathbf{A}\mathbf{x} - \mathbf{b} = 0 \\
\mathbf{x} \geq & \quad 0
\end{aligned}
$$

où $\mathbf{e}$ est un vecteur de 1 et $\mathbf{X}^{-1}$ est la matrice diagonale des $\frac{1}{x_i}$. On a des conditions de stationnarité qui sont non-linéaires.

Réécriture

- On définit $\mathbf{s} = \mu \mathbf{X}^{-1}\mathbf{e}$
- On obtient le système d'équation suivant

$$
\begin{aligned}
\Rightarrow \mathbf{Q}\mathbf{x} - \mathbf{s} - \mathbf{A}^\top \lambda &= -c \\
\mathbf{A}\mathbf{x} &= \mathbf{b} \\
x_i s_i &= \mu \quad \forall i = 1, \cdots, d
\end{aligned}
$$

QP méthode de points intérieurs

KKT et objectifs

- pour un $\mu > 0$ donné, les conditions KKT définissent un système d'équations non-linéaires.
- Soit
$$F_\mu(\mathbf{x}, \lambda, \mathbf{s}) = \begin{bmatrix} \mathbf{Qx} + \mathbf{c} - \mathbf{s} - \mathbf{A}^\top \lambda \\ \mathbf{Ax} - \mathbf{b} \\ XS - \mu \mathbf{e} \end{bmatrix}$$
où $X = diag(\mathbf{x}) \in \mathbb{R}^{d \times d}$ et $S = diag(\mathbf{s}) \in \mathbb{R}^{d \times d}$.
- les conditions KKT sont satisfaites si on trouve $\mathbf{x}, \lambda$ et $\mathbf{s}$ telles que
$$F_\mu(\mathbf{x}, \lambda, \mathbf{s}) = 0$$

QP méthode de points intérieurs

Problème sous-jacent

- Trouver les zeros d'une fonction multivariée

$$F_\mu(\mathbf{x}, \lambda, \mathbf{s}) = 0$$

Méthode de Newton-Raphson

- en 1D : soit x_k une approximation du 0 alors, x_{k+1}

$$0 = f'(x_k)(x_{k+1} - x_k) + f(x_k)$$

  est une meilleure approximation
- en multivarié, on cherche $\mathbf{d}$ telle que

$$0 = F_\mu(\mathbf{x}_k, \lambda_k, \mathbf{s}_k) + J_{F_\mu}(\mathbf{x}_k, \lambda_k, \mathbf{s}_k)\mathbf{d}_k$$

  où J est la jacobienne de F et $\mathbf{d}_k$ la direction de mise à jour.
  $\mathbf{d}_k = (\Delta\mathbf{x}_k, \Delta\lambda_k, \Delta\mathbf{s}_k)$ et
  $(\mathbf{x}_{k+1}, \lambda_{k+1}, \mathbf{s}_{k+1}) = (\mathbf{x}_k, \lambda_k, \mathbf{s}_k) + \alpha_k(\Delta\mathbf{x}_k, \Delta\lambda_k, \Delta\mathbf{s}_k)$

QP méthode de points intérieurs

Itération d'une méthode de point intérieur

- résolution d'un système linéaire en d défini par

$$-F_\mu(\mathbf{x}_k, \lambda_k, \mathbf{s}_k) = J_{F_\mu}(\mathbf{x}_k, \lambda_k, \mathbf{s}_k)\mathbf{d}$$

- La Jacobienne s'écrit

$$J_{F_\mu}(\mathbf{x}_k, \lambda_k, \mathbf{s}_k) = \begin{bmatrix} \mathbf{Q} & -\mathbf{A}^\top & -\mathbf{I} \\ \mathbf{A} & 0 & \mathbf{0} \\ S_k & 0 & X_k \end{bmatrix}$$

  où $X_k = diag(\mathbf{x}_k) \in \mathbb{R}^{d\times d}$ et $S_k = diag(\mathbf{s}_k) \in \mathbb{R}^{d\times d}$.
- La résolution de ce système linéaire est couteux, il faut un algorithme efficace et approprié.

QP méthode de points intérieurs

Algorithme

1. initialiser l'algorithme avec des paramètres $\mathbf{x}_0 > 0$, $\mathbf{s}_0 > 0, k = 0$
2. définir $\sigma_k \in [\sigma_{min}, \sigma_{max}]$, $\mu_k = \frac{\mathbf{x}_k^\top \mathbf{s}_k}{d}$
3. résoudre
$$J_{F_\mu}(\mathbf{x}_k, \lambda_k, \mathbf{s}_k)\mathbf{d} = - \begin{bmatrix} \mathbf{Q}\mathbf{x} + \mathbf{c} - \mathbf{s} - \mathbf{A}^\top \lambda \\ \mathbf{A}\mathbf{x} - \mathbf{b} \\ XS - \sigma_k \mu_k \mathbf{e} \end{bmatrix}$$
pour obtenir $\mathbf{d}_k$
4. en partant de $\alpha = 0$, choisir le plus grand α_k tel que
$$(\mathbf{x}_k, \mathbf{s}_k) + \alpha_k(\Delta\mathbf{x}_k, \Delta\mathbf{s}_k) > 0$$
5. mise à jour
$$(\mathbf{x}_{k+1}, \lambda_{k+1}, \mathbf{s}_{k+1}) = (\mathbf{x}_k, \lambda_k, \mathbf{s}_k) + \alpha_k(\Delta\mathbf{x}_k, \Delta\lambda_k, \Delta\mathbf{s}_k)$$
6. $k = k + 1$
7. si $\frac{\mathbf{x}_k^\top \mathbf{s}_k}{d} \leq \epsilon$ retourner en 2

QP méthode de points intérieurs

Remarques

- critère de terminaison $\frac{\mathbf{x}_k^\top \mathbf{s}_k}{d} \leq \epsilon$
  - comme $\mathbf{x}^\top \mathbf{s} = d\mu$, le critère d'arret suppose que la barrière a de moins en moins d'influence.
- le rôle de σ_k
  - idéalement, on cherche un zéro de F_μ quand $\mu = 0$.
  - on part d'un μ donné, et on décroit le μ cible en prenant $\sigma_k \mu_k$ et $\sigma_k < 1$
  - typiquement, on prend $\sigma_k \in [0.01, 0.75]$.
- la structure de la Jacobienne peut etre prise en compte pour construire un solveur de système linéaire efficace
- il faut trouver un point initial $\mathbf{x}_0, \mathbf{s}_0$ faisable