Master 2 informatique

Parcours IAAA, UE MIA

Année 2022-23

DM 3 : Optimisation et Apprentissage

Exercice 1 Conditions de Karush-Kuhn-Tucker pour des fonctions $\mathcal{C}^1$.

Soit f, $(e_i)_{i=1}^{n_E}$ et $(c_i)_{i=1}^{n_I}$ des fonctions de $\mathbf{R}^d$ vers $\mathbf{R}$ de classe $\mathcal{C}^1$.

Comme nous l'avons vu en cours, on dit que le point $x^* \in \mathbf{R}^d$ vérifie les conditions de Karush-Kuhn-Tucker (KKT) pour le problème :

$$\min_{x \in \mathbf{R}^d} f(x) \text{ tel que } \begin{cases} e_i(x) = 0, & i \in \mathcal{E} \\ c_i(x) \geq 0, & i \in \mathcal{I} \end{cases}$$

si et seulement si il existe $\lambda_E^* \in \mathbf{R}^{n_E}$ et $\lambda_I^* \in \mathbf{R}^{n_I}$, tels que :

1. $\nabla_x \mathcal{L}(x^*, \lambda_E^*, \lambda_I^*) = 0$, où $\mathcal{L}(x, \lambda_E, \lambda_I) = f(x) - \sum_{i=1}^{n_E} \lambda_{E,i} e_i(x) - \sum_{i=1}^{n_I} \lambda_{I,i} c_i(x)$
2. $e_i(x^*) = 0$ pour tout i dans $\{1, 2, \ldots, n_E\}$
3. $c_i(x^*) \geq 0$ pour tout i dans $\{1, 2, \ldots, n_I\}$
4. $\lambda_{I,i} \geq 0$ pour tout i dans $\{1, 2, \ldots, n_I\}$
5. $\lambda_{I,i} c_i(x^*) = 0$ pour tout i dans $\{1, 2, \ldots, n_I\}$

Les trois théorèmes suivants donnent des conditions de régularité sur les fonctions du problème—appelées *qualifications des contraintes*—qui garantissent que toute solution locale du problème considéré doit vérifier les conditions KKT (c'est à dire que sous ces conditions de régularité les conditions KKT sont des conditions nécessaires de solution locale). Notez qu'on peut démontrer la validité d'autres qualification des contraintes que les trois présentées ici et qu'il est possible de généraliser ces théorèmes au delà du cas des fonctions de classe $\mathcal{C}^1$.

Théorème. (Conditions nécessaires de solution locale pour des contraintes linéaires.)

Si les fonctions e_i et c_j sont affines pour $i = 1 \ldots n_E$ et $j = 1 \ldots n_I$ alors les conditions KKT sont des conditions nécessaires de solution locale pour le problème considéré.

Théorème. (Conditions nécessaires de solution locale pour des contraintes linéairement indépendantes.)

Si au point $x \in \mathbf{R}^d$ la matrice formée par les gradients des contraintes d'égalité et des contraintes d'inégalité **actives** est de rang plein (c'est à dire que les n_E vecteurs gradients pour les contraintes d'égalité—qui doivent toujours toutes être actives—et les vecteurs gradients pour les contraintes d'inégalité actives en x—c'est à dire telles que $c_i(x) = 0$—sont linéairement indépendants), alors les conditions KKT sont des conditions nécessaires de solution locale en x pour le problème considéré.

Théorème. (Conditions nécessaires de solution locale pour un problème convexe.)
Si f est convexe, c_i est concave pour tout i in $1 \dots n_I$ et e_j est affine pour tout j in $1 \dots n_E$ et si la condition de Slater est vérifiée, c'est à dire s'il existe $x \in \mathbf{R}^d$ tel que $e_j(x) = 0$ pour tout j in $1 \dots n_E$ et $c_i(x) > 0$ pour tout i in $1 \dots n_I$, alors les conditions KKT sont des conditions nécessaires de solution locale pour le problème considéré.

Les trois questions de cet exercice sont indépendantes.

1. On considère le problème

$$\min_{x,y \in \mathbf{R}^2} x^2 + y^2 \text{ tel que } \begin{cases} x + y \geq 1 \\ y \leq 2 \\ y^2 \geq x \end{cases}$$

   (a) Donnez les conditions KKT pour ce problème.

   (b) Trouvez tous les points solutions des conditions de KKT pour ce problème.

   (c) Donnez en le justifiant toutes les solutions globales de ce problème.

   (d) Faites un graphe pour vérifier la plausibilité de vos réponses.

2. Trouvez le parallélépipède rectangle (pavé droit) dont la surface (somme des aires des six faces) est maximale parmi tous les parallélépides rectangles de diagonale $x_1^2 + x_2^2 + x_3^2 = L$, où L est un nombre réel strictement positif fixé et x_1, x_2, x_3 sont les longueurs des côtés du parallélépipède rectangle.

3. Soit L, L' et λ trois réels strictement positifs, y une matrice colonne de taille n et A une matrice de taille n par m. On considère les trois problèmes d'optimisation suivants :

$$\min \|y - Ax\|_2, \text{ pour } x \in \mathbf{R}^m \text{ tel que } \|x\|_2 \leq L. \tag{1}$$

$$\min \|y - Ax\|_2^2, \text{ pour } x \in \mathbf{R}^m \text{ tel que } \|x\|_2^2 \leq L'. \tag{2}$$

$$\min \|y - Ax\|_2^2 + \lambda \|x\|_2^2, \text{ pour } x \in \mathbf{R}^m. \tag{3}$$

   Montrez que pour tout choix de $L > 0$, il existe un choix de $L' > 0$ et $\lambda > 0$ tels que les solutions globales des trois problèmes soient identiques.

Exercice 2 Régression linéaire

On observe $(x_1, y_1), \dots (x_n, y_n) \in (\mathbf{R}^d \times \mathbf{R})^n$. On modélise $x_1, \dots, x_n$ comme les valeurs observées de variables aléatoires $X_1, \dots, X_n$ i.i.d. de loi P et $y_1, \dots, y_n$ comme les valeurs observées de variables aléatoires $Y_1, \dots Y_n$ telles qu'il existe $\theta^* \in \mathbf{R}^d$ tel que pour tout i dans $\{1, \dots n\}$, $Y_i = X_i^T \theta^* + \epsilon_i$, avec $\epsilon_1, \dots, \epsilon_n$ i.i.d. de loi N telle que $\mathbf{E}(N) = 0$ et $\mathbf{Var}(N) = \sigma^2$.

On note X la matrice de taille n par d dont les lignes sont $x_1, \dots, x_n$ et Y la matrice colonne contenant $y_1, \dots, y_n$. On cherche à estimer θ^* et on considère deux estimateurs :

— L'estimateur de la régression « ridge » :

$$\hat{\theta}_\lambda := \arg \min_{\theta \in \mathbf{R}^d} \frac{1}{n} \|Y - X\theta\|_2^2 + \lambda \|\theta\|_2^2,$$

défini pour $\lambda > 0$;

— L'estimateur de norme minimale parmi les estimateurs au moindre carrés :

$$\hat{\theta} := \arg \min_{\theta \in \Theta_{LS}} \|\theta\|_2$$

où $\Theta_{LS} = \arg\min_{\theta \in \mathbf{R}^d} \|Y - X\theta\|_2^2$.

1. Montrez que pour toute matrice M de taille n par m et tout réel $\delta > 0$, $M^T M + \delta I$ est inversible. (Vous pouvez montrer, par exemple, que la matrice est symmétrique définie positive et utiliser le théorème spectral pour conclure.)
2. Montrez que le problème d'optimisation dont $\hat{\theta}_\lambda$ est défini comme la solution possède bien une unique solution donnée par $\hat{\theta}_\lambda = \frac{1}{n} \left(\frac{X^T X}{n} + \lambda I \right)^{-1} X^T Y$.
3. Montrez que le problème d'optimisation dont $\hat{\theta}$ est défini comme la solution possède bien une unique solution donnée par $\hat{\theta} = X^\dagger Y$, où $X^\dagger$ est la pseudo-inverse (de Moore-Penrose) de X.
4. Montrez que $\hat{\theta}_\lambda$ tend vers $\hat{\theta}$ quand λ tend vers 0 par valeurs positives. (Indice : commencer par le cas où $d = n = 1$, puis le cas où X est diagonale, puis le cas général.)
5. (Bonus) Montrer qu'une simple descente de gradient à pas fixe sur $f(\theta) := \|Y - X\theta\|^2$ converge vers $\hat{\theta}$. Précisez comment choisir le pas. (Vous pouvez vous inspirer de l'exercice sur la convergence de la descente de gradient dans le DM2.)
6. Calculez le biais de $\hat{\theta}_\lambda$.
7. Calculez la variance de $\hat{\theta}_\lambda$.

8. (Bonus) Calculez le biais et la variance de $\hat{\theta}$ et commentez.