Introduction aux fonctions de décision non-linéaires

Thierry Bazier-Matte

13 mai 2017

1 Algorithme de décision optimale

1.1 Formulations primale et duale

Tel que discuté en introduction, le cas le plus simple pour un espace de décision est celui où l'espace $\boldsymbol{X}$ des variables de marché ne subit aucune transformation. L'espace $\boldsymbol{Q}$ correspond [1] alors à $\boldsymbol{X}$ et le scalaire de décision $q(x)$ est obtenu par produit scalaire $q^T x$. On obtient donc ici le problème sous la forme *primale* :

$$\boxed{\underset{q \in \boldsymbol{X}}{\text{maximiser}} \quad n^{-1} \sum_{i=1}^{n} u(r_i\, q^T x_i) - \frac{\lambda}{2} \|q\|^2} \tag{1}$$

Formulation primale I

Par la théorie de l'optimisation convexe (voir [**?**]) on sait qu'une solution $\hat{q}$ existe et qu'elle est unique. En supposant que $\boldsymbol{X} \subseteq \mathscr{R}^p$, on peut alors exprimer $\hat{q}$ comme une combinaison linéaire de p coordonnées.

Cependant, $\hat{q}$ peut aussi être exprimé comme une combinaison linéaires de la base $\{x_1, \ldots, x_n\}$, *i.e.* il existe un vecteur $\hat{\alpha} \in \mathscr{R}^n$ tel que $\hat{q} = \Xi^T \hat{\alpha}$, où $\Xi \in \mathscr{R}^{n \times p}$ est la matrice des observations.

Il suffit en effet de remarquer que l'espace $\boldsymbol{Q} = \boldsymbol{X}$ peut être décomposé comme la somme directe du sous-espace vectoriel $\hat{\boldsymbol{X}}$ engendré par Ξ^T et son complément orthogonal $\hat{\boldsymbol{X}}^\perp$, *i.e.* $\boldsymbol{Q} = \hat{\boldsymbol{X}} \oplus \hat{\boldsymbol{X}}^\perp$. Ainsi, tout vecteur de décision $q \in \boldsymbol{Q}$ s'exprime comme la somme de $\hat{x} \in \hat{\boldsymbol{X}}$ et $\hat{x}^\perp \in \hat{\boldsymbol{X}}^\perp$, deux vecteurs orthogonaux (voir les appendices de [**?**] ou [MRT12]). La fonction objectif $\boldsymbol{EU}_\lambda$ évaluée au point q se simplifie alors ainsi :

$$\boldsymbol{EU}_\lambda(q) = \boldsymbol{EU}_\lambda(\hat{x} + \hat{x}^\perp) \tag{2}$$

1. En fait, pour être exact, $q : \boldsymbol{X} \to \mathscr{R}$ étant une *fonction*, il est plus exact de faire correspondre $\boldsymbol{Q}$ à l'espace *dual* $\boldsymbol{X}^*$. Mais par le théorème de Riez, à tout vecteur $v \in \boldsymbol{V}$ d'un espace vectoriel, il existe un unique vecteur dual $v^* \in \boldsymbol{V}^*$ (noté v^T en dimension finie). On peut donc se contenter de faire correspondre $\boldsymbol{Q}$ à $\boldsymbol{X}$.

$$= n^{-1} \sum_{i=1}^{n} u(r_i\,(\hat{x} + \hat{x}^{\perp})^T x_i) - \frac{\lambda}{2} \|\hat{x} + \hat{x}^{\perp}\|^2 \tag{3}$$

$$\leq n^{-1} \sum_{i=1}^{n} u(r_i\,\hat{x}^T x_i) - \frac{\lambda}{2} \|\hat{x}\|^2, \tag{4}$$

puisque, par définition, $(\hat{x}^{\perp})^T x_i = 0$ pour toute observation x_i et que d'autre part $\|\hat{x}\|^2 \leq \|\hat{x}\|^2 + \|\hat{x}^{\perp}\|^2 = \|q\|^2$. Ainsi, toute solution $\hat{q}$ repose bien dans l'espace colonne de Ξ^T.

Cette observation (qui correspond en fait au célèbre théorème de la représentation) peut se révéler très utile car elle permet de changer le domaine d'optimisation de $\boldsymbol{X}$ à $\mathscr{R}^n$ par l'identité $\hat{q} = \Xi^T \hat{\alpha}$. Autrement dit, le problème d'optimisation devient

$$\underset{\alpha \in \mathscr{R}^n}{\text{maximiser}} \quad n^{-1} \sum_{i=1}^{n} u(r_i\,\alpha^T \Xi x_i) - \frac{\lambda}{2} \alpha^T \Xi \Xi^T \alpha. \tag{5}$$

On peut simplifier cette expression en posant $K \coloneqq \Xi\Xi^T \in \mathscr{R}^{n \times n}$:

$$\boxed{\underset{\alpha \in \mathscr{R}^n}{\text{maximiser}} \quad n^{-1} \sum_{i=1}^{n} u(r_i K_i \alpha) - \frac{\lambda}{2} \alpha^T K \alpha,} \tag{6}$$

Formulation duale I

où $\sum_{j=1}^{n} x_j^T x_i = K_i \in \mathscr{R}^n$ représente la i^{e} colonne (ou rangée car K est alors symétrique) de K. En fait K correspond à la *matrice gramienne*, *i.e.* la matrice des produits scalaires de toutes les observations x_i.

Le problème dual, lorsqu'il est exprimé avec une utilité risque neutre, admet une solution analytique équivalente à la solution primale, mais qui donne une interprétation supplémentaire au concept de solution optimale. En considérant

$$\underset{\alpha \in \mathscr{R}^n}{\text{maximiser}} \quad n^{-1} r^T K \alpha - \frac{\lambda}{2} \alpha^T K \alpha \tag{7}$$

et en faisant l'hypothèse qu'aucune observation x_i n'est de norme nulle (une telle observation serait de toute façon inutile puisqu'elle n'induirait alors aucune décision d'investissement, étant donné que $q^T x_i = 0$ pour toute décision q), K est alors définie positive et la solution $\hat{\alpha} = (\lambda n)^{-1} r$ est unique. Ainsi, la décision optimale $\hat{q} = \Xi^T \hat{\alpha}$ peut être conçue comme une moyenne pondérée par le rendement de chacune des observations.

1.2 Transformations non linéaires

Le cas $\boldsymbol{Q} = \boldsymbol{X}$ est cependant trop pour rendre compte de certaines géométries de problème. Il est alors naturel de définir une transformation non linéaire d'une observation $x \in \boldsymbol{X}$ par une fonction $\phi : \boldsymbol{X} \to \phi(\boldsymbol{X})$. Par exemple, si $\boldsymbol{X} \in \mathscr{R}$, *i.e.* une seule variable de marché est considérée, alors on peut chercher une solution polynômiale en posant une transformation du genre

$$\phi : x \mapsto \begin{pmatrix} 1 \\ x \\ x^2 \\ \vdots \\ x^k \end{pmatrix}. \tag{8}$$

En notant $\langle \cdot, \cdot \rangle$ le produit scalaire de l'espace $\phi(\boldsymbol{X})$, le problème consiste alors à trouver un vecteur optimal $q \in \boldsymbol{Q} = \phi(\boldsymbol{X})$ de façon à

$$\boxed{\underset{q \in \phi(\boldsymbol{X})}{\text{maximiser}} \quad n^{-1} \sum_{i=1}^{n} u(r_i \langle q, \phi(x_i) \rangle) - \frac{\lambda}{2} \|q\|^2.} \tag{9}$$

Formulation primale II

Mais puisque le théorème de représentation s'applique encore, $\hat{q}$ peut aussi s'exprimer comme une combinaison linéaire α des observations $\{\phi(x_1), \ldots, \phi(x_n)\}$:

$$\boxed{\underset{\alpha \in \mathscr{R}^n}{\text{maximiser}} \quad n^{-1} \sum_{i=1}^{n} u(r_i K_i \alpha) - \frac{\lambda}{2} \alpha^T K \alpha.} \tag{10}$$

Formulation duale II

Cette fois par contre $K_{ij} = \langle \phi(x_i), \phi(x_j) \rangle$; chaque élément de K représente le produit scalaire dans le nouvel espace obtenu par transformation ϕ.

1.3 Fonctions de noyau

Ainsi, pour toute transformation $\phi : \boldsymbol{X} \to \phi(\boldsymbol{X})$, quelle que soit la dimension de l'espace $\phi(\boldsymbol{X})$, on peut exprimer la décision optimale à partir d'une optimisation sur n dimensions. En outre, ce programme d'optimisation ne dépend plus que du produit scalaire entre ces observations transformées. De plus, on peut dans bien des cas court-circuiter le calcul de ce produit scalaire par une fonction *noyau* $\kappa : \boldsymbol{X} \times \boldsymbol{X} \to \mathscr{R}$ telle que $\kappa(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$.

Par exemple, dans l'exemple montré plus haut, il suffirait de poser

$$\kappa(x_i, x_j) = 1 + x_i x_j + (x_i x_j)^2 + \cdots + (x_i x_j)^k. \tag{11}$$

Évidemment, dans un pareil cas le gain est assez faible puisqu'on a uniquement réarrangé l'ordre des opérations. Mais, il est alors possible de circonvenir complètement la transformation ϕ et de représenter sa non linéarité qu'à partir de κ. Par exemple, le noyau gaussien

$$\kappa(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \tag{12}$$

permet de calculer directement le produit scalaire $\langle \phi(x_i), \phi(x_j) \rangle$ d'observations $\phi(x_i)$ et $\phi(x_j)$ de dimension infine.

Choisir adéquatement le noyau est alors une tâche cruciale du modèle puisque tout noyau κ induit une géométrie particulière du modèle ; il peut alors être impossible de déterminer une fonction de décision q pourvue de bonne performance si le noyau ne correspond pas à la géométrie de la loi de marché M.

Il faut par ailleurs imposer une contrainte supplémentaire à la classe des noyaux possibles. En effet $\kappa(x_i, x_j)$ représente un produit scalaire dans $\phi(\boldsymbol{X})$ et est donc tenu de respecter les popriétés algébriques de celui-ci. En fait, κ doit satisfaire l'inégalité de Cauchy-Schwartz :

$$\begin{aligned} 0 \leq \kappa(x_i, x_j)^2 &= \langle \phi(x_i), \phi(x_j) \rangle^2 && (13) \\ &\leq \|\phi(x_i)\| \|\phi(x_j)\| && (14) \\ &= \kappa(x_i, x_i) \kappa(x_j, x_j). && (15) \end{aligned}$$

Le noyau κ doit donc être une forme semi-définie positive sur $\boldsymbol{X}$ (voir encore [MRT12]).

Références

[KW71] George Kimeldorf and Grace Wahba. Some results on tchebycheffian spline functions. *Journal of mathematical analysis and applications*, 33(1) :82–95, 1971.

[MRT12] Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of machine learning*. MIT press, 2012.