TD 03 : Premiers pas avec le modèle linéaire

Fondements du Machine Learning, L3 IM2D

Octobre 2024

Exercice 3.1 : Modélisation par moindres carrés

On dispose de m notes $y_1, \ldots, y_m$ données par des individus ayant chacun passé une nuit dans le même hôtel. On cherche à déterminer une note globale de l'hôtel qui soit la plus cohérente possible avec l'ensemble des m notes.

a) Modéliser le problème sous la forme d'un système linéaire et d'un problème aux moindres carrés.

b) Pour les deux formulations obtenues, décrire si le problème possède ou non une solution, et donner l'ensemble des solutions s'il existe.

c) On suppose maintenant que chaque individu i a passé x_i nuits dans l'hôtel. Comment peut-on tenir compte de cela dans la modélisation ? Quel serait alors l'impact sur la résolution des problèmes associés ?

Exercice 3.2 : Matrices de rang 1

Soit une matrice $\boldsymbol{X} \in \mathbb{R}^{m \times 1}$ de rang 1.

a) Donner une décomposition en valeurs singulières de $\boldsymbol{X}$.

b) Soit $\boldsymbol{y} \in \mathbb{R}^m$. Résoudre le problème aux moindres carrés

$$\min_{\beta \in \mathbb{R}} \frac{1}{2} \left\| \boldsymbol{X}\beta - \boldsymbol{y} \right\|^2 .$$

c) On suppose que $\boldsymbol{X}$ est une matrice orthogonale. Que représente alors $\boldsymbol{X}\beta^*$, où β^* est une solution du problème aux moindres carrés ?

Exercice 3.3 : Moindres carrés régularisés

(Adapté d'un exercice d'examen proposé en 2019-2020.)

Étant donnés $\boldsymbol{X} \in \mathbb{R}^{m \times n}$ et $\boldsymbol{y} \in \mathbb{R}^m$, on considère donc le problème aux moindres carrés suivant :

$$\min_{\beta \in \mathbb{R}^n} \frac{1}{2}\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2 + \frac{\lambda}{2}\|\boldsymbol{\beta}\|_2^2, \tag{1}$$

avec $\lambda \geq 0$ (on englobe ainsi le cas non régularisé vu en cours, qui correspond à $\lambda = 0$).

a) On suppose dans cette question que $\boldsymbol{X}$ est de rang n avec $n < m$.

   i) Quelle propriété sur $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ cela implique-t-il ?
   En déduire la solution de (1) lorsque $\lambda = 0$.

   ii) Lorsque $\lambda > 0$, $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} + \lambda\boldsymbol{I}$ est définie positive. On peut alors montrer que les solutions du problème (1) sont exactement celles du problème

   $$\min_{\boldsymbol{\beta} \in \mathbb{R}^n} \frac{1}{2}\|(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} + \lambda\boldsymbol{I})^{1/2}\boldsymbol{\beta} - \boldsymbol{z}\|_2^2,$$

   où $\boldsymbol{I}$ est la matrice identité de $\mathbb{R}^{n \times n}$, et $\boldsymbol{z}$ vérifie $(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} + \lambda\boldsymbol{I})^{1/2}\boldsymbol{z} = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{y}$. [1]
   En déduire l'ensemble des solutions du problème (1) dans ce cas.

b) On suppose maintenant que $\boldsymbol{X}$ est de rang $m < n$.

   i) Une solution au problème (1) dans le cas $\lambda = 0$ est donné par $\boldsymbol{X}^{\dagger}\boldsymbol{y}$, où $\boldsymbol{X}^{\dagger} = \boldsymbol{X}^{\mathrm{T}}(\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}})^{-1}$ est la pseudo-inverse de $\boldsymbol{X}$. Quelle propriété possède cette solution?

   ii) Lorsque $\lambda > 0$, montrer en utilisant la décomposition en valeurs singulières de $\boldsymbol{X}$ que la matrice $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} + \lambda\boldsymbol{I}$ est de rang plein.

   iii) Comment caractériser la ou les solution(s) du problème (1) en utilisant le résultat des questions a)ii) et b)ii) ?

[1] On rappelle la notion de racine carrée de matrice déjà utilisée dans le TD 1. Si $\boldsymbol{A} \succ \boldsymbol{0}$, alors $\boldsymbol{A}^{1/2}$ est l'unique matrice symétrique définie positive telle que $(\boldsymbol{A}^{1/2})^2 = \boldsymbol{A}$.