Exemples de problèmes d'optimisation en apprentissage automatique

Quadratique convexe

Dérivable convexe

Convexe non-dérivable

Convexe avec contraintes

Non-convexe dérivable

Non-convexe non dérivable

Combinatoire

Moindres carrés

Modèle linéaire : $y_i = x_i^\top \beta + \varepsilon_i$
x_i : variables explicatives pour l'observation i
β : vecteur des paramètres du modèle
ε_i : bruit

Problème d'optimisation

$$\min_{\beta \in \mathbb{R}^p} \frac{1}{2} \|X\beta - y\|_2^2 = \min_{\beta \in \mathbb{R}^p} \frac{1}{2} \sum_{i=1}^{n} (x_i^\top \beta - y_i)^2$$

$X \in \mathbb{R}^{n \times p}, y \in \mathbb{R}^n$

Retour accueil

Régularisation de Tikhonov (ridge regression)

Modèle linéaire : $y_i = x_i^\top \beta + \varepsilon_i$
On veut forcer β à ne pas avoir des coefficients immenses

Problème d'optimisation

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

$X \in \mathbb{R}^{n \times p}, y \in \mathbb{R}^n, \lambda > 0$

Retour accueil

Régression logistique

Modèle linéaire : $y_i = x_i^\top w + w_0 + \varepsilon_i$
Classification : $y_i \in \{-1, 1\}$
Classifieur :
$h : x \mapsto \mathrm{sign}(\langle x, w\rangle + w_0) \qquad (w \in \mathbb{R}^p, w_0 \in \mathbb{R})$

Problème d'optimisation

$$\min_{w\in\mathbb{R}^p, w_0\in\mathbb{R}} \frac{1}{n}\sum_{i=1}^{n} \log(1+\exp(-y_i(x_i^\top w + w_0))) + \frac{\lambda}{2}\|w\|^2$$

Retour accueil

Lasso

Modèle linéaire : $y_i = x_i^\top \beta + \varepsilon_i$
On veut un vecteur de paramètre β parcimonieux (beaucoup de coefficients égaux à 0)

Problème d'optimisation

$$\min_{\beta \in \mathbb{R}^p} \frac{1}{2} \|X\beta - y\|_2^2 + \lambda \|\beta\|_1$$

$X \in \mathbb{R}^{n \times p}, y \in \mathbb{R}^n, \lambda > 0$
$\|\beta\|_1 = \sum_{i=1}^p |\beta_i|$ pas dérivable mais séparable

Retour accueil

Séparateurs à Vaste Marge (SVM)

Modèle linéaire : $y_i = x_i^\top w + w_0 + \varepsilon_i$
Classification : $y_i \in \{-1, 1\}$
Classifieur :
$h : x \mapsto \text{sign}(\langle x, w \rangle + w_0) \qquad (w \in \mathbb{R}^p, w_0 \in \mathbb{R})$

Problème d'optimisation

$$\min_{w \in \mathbb{R}^p, w_0 \in \mathbb{R}} \sum_{i=1}^{n} C \max(0, 1 - y_i(x_i^\top w_i + w_0)) + \frac{1}{2} \|w\|^2$$

Retour accueil

SVM dual

Modèle linéaire : $y_i = x_i^\top w + w_0 + \varepsilon_i$
Classification : $y_i \in \{-1, 1\}$

Problème d'optimisation

$$\min_{\alpha \in \mathbb{R}^n} \sum_{j=1}^{p} \left(\sum_{i=1}^{n} \alpha_i y_i X_{i,j}\right)^2 - \sum_{i=1}^{n} \alpha_i$$

$$0 \leq \alpha_i \leq C, \forall i, \qquad \sum_{i=1}^{n} y_i \alpha_i = 0$$

$w_j = \sum_{i=1}^{n} \alpha_i y_i X_{i,j}$

Retour accueil

Réseaux de neurones

Modèle non-linéaire : $y_i = f(x_i, w) + \varepsilon_i$

Exemple : 1 couche cachée avec H neurones

$$f(x,w) = \sigma\left(\sum_{i=1}^{H} w_i v_i\left(\sum_{i=j}^{p} w_{i,j} x_j\right)\right)$$

$\sigma, v_1, \ldots, v_H$ sont appelées fonctions d'activation

Problème d'optimisation

$$\min_{w \in \mathbb{R}^{H+Hp}} \sum_{i=1}^{n} \|y_i - f(x_i, w)\|^2$$

Retour accueil

ACP et NMF

A matrice de taille $n \times p$

ACP :

$$\min_{U \in \mathbb{R}^{n \times k}, V \in \mathbb{R}^{k \times p}} \|A - UV\|_F^2$$

NMF :

$$\min_{U \in \mathbb{R}^{n \times k}, V \in \mathbb{R}^{k \times p}} \|A - UV\|_F^2$$

$$U_{i,j} \geq 0, V_{j,l} \geq 0$$

Retour accueil

0-1 Loss

Modèle linéaire : $y_i = x_i^\top w + w_0 + \varepsilon_i$
Classification : $y_i \in \{-1, 1\}$
Classifieur :
$h : x \mapsto \mathrm{sign}(\langle x, w\rangle + w_0) \qquad (w \in \mathbb{R}^p, w_0 \in \mathbb{R})$

Problème d'optimisation

$$\min_{w \in \mathbb{R}^p, w_0 \in \mathbb{R}} \sum_{i=1}^{n} \mathbb{1}(y_i(x_i^\top w + w_0) < 0)$$

Retour accueil

K plus proches voisins

x_0 fixé, on cherche les K plus proches voisin de x_0

$$\min_{s \in \mathbb{Z}^n} \sum_{i=1}^{n} s_i \|x_0 - x_i\|$$
$$0 \leq s_i \leq 1, \forall i$$
$$\sum_{i=1}^{n} s_i = K$$

Retour accueil

K-means

On cherche à partitionner les points $(x_j)_{1 \leq j \leq n}$ en K ensembles $S_1, \ldots S_K$

$$\min_{S_1,\ldots,S_K} \sum_{k=1}^{K} \sum_{x_j \in S_i} \|x_j - \mu_k\|^2$$

$$\mu_k = \frac{1}{|S_k|} \sum_{x_j \in S_i} x_j, \forall k$$

$$S_j \cap S_i = \emptyset, \forall i, j$$

$$S_1 \cup \ldots S_K = \{1, \ldots, n\}$$

Retour accueil