Identification de paramètres et optimisation
Cours de Master 2 STIM
2014-2015

Sébastien Adam

3 décembre 2015

Présentation (1)

Qui suis-je ?

- Sébastien ADAM
  - Professeur, laboratoire LITIS
  - Équipe document et apprentissage
  - Sebastien.Adam@univ-rouen.fr
  - 02.32.95.52.10
  - Bureau U2.1.41
  - Responsable de la spécialité GEII du Master IGIS

Équipe « DocApp »

- 18 enseignants chercheurs + 20 doctorants
- Apprentissage statistique, séquences, structurel, ensembles
- Application : analyse d'images de document, BCI, RI

Présentation (2)

Mes thématiques de recherche

- Reconnaissance de formes - apprentissage
  - Forêts aléatoires
  - Apprentissage multi-objectif
- Représentations structurelles en reconnaissance de formes
  - Distances entre graphes
  - Recherche d'isomorphismes de sous graphes
  - Classification de graphes

Applications cibles

- Analyse d'images de documents
- Analyse d'images médicales

Présentation (3)

Unité d'Enseignement Optimisation et apprentissage

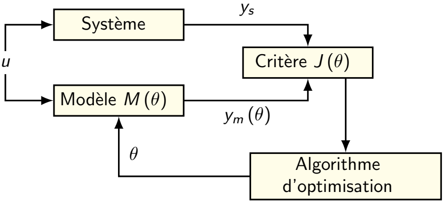
- Objectifs :
  - Maîtriser certains outils de base utilisés en optimisation (Moindres carrés, Gradient, Gauss-Newton, Recuit Simulé)
  - Introduction aux réseaux de neurones
  - Introduction aux méthodes évolutionnaires
  - Introduction à l'optimisation multi-objectifs, ROC-based learning
  - Liens optimisation/Apprentissage
- Volumes horaires :
  - 15h de Cours
  - 15h de TP
- Evaluation : 60 points
  - Un examen terminal écrit : 40 points
  - Une note de TP : 20 points
- Cours disponible sur universitice

Plan du Cours

Introduction

Cadre historique du cours : la modélisation paramétrique

- Objectif : identifier les paramètres θ du modèle d'un système pour que son comportement « ressemble » au maximum à celui du système..

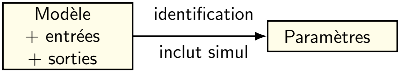


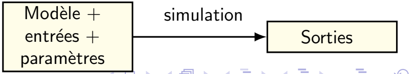
- K paramètres
- N observations du système

$$\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_K \end{pmatrix} \quad y_S = \begin{pmatrix} Y_{S1} \\ Y_{S2} \\ \vdots \\ Y_{SN} \end{pmatrix} \quad y_M = \begin{pmatrix} y_{M1} \\ y_{M2} \\ \vdots \\ y_{MN} \end{pmatrix}$$

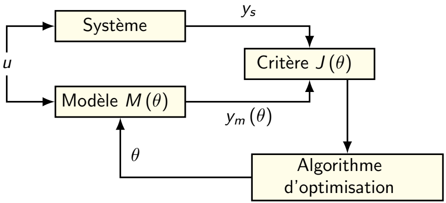
Objectifs de la modélisation

- Centré « modèle » : comprendre le fonctionnement d'un système
- Centré « commande » : utiliser le modèle pour piloter le système





Introduction



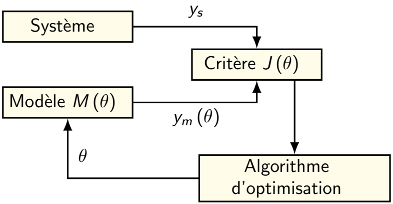
Les problèmes à résoudre en modélisation paramétrique

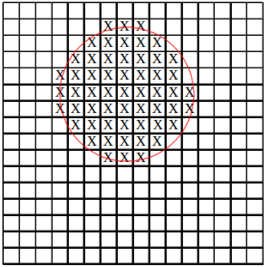
- Quel modèle ?
  - Statique / Dynamique ?
  - Continu / Discret ?
  - Quel niveau de connaissances : loi générales / boîtes noires / boîtes grises ?
- Quel critère pour quantifier la qualité d'un modèle ?
- Quelle méthode pour déterminer (optimiser) la valeur des paramètres ?

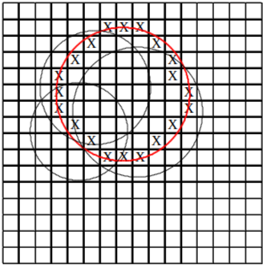
Introduction

Exemple 1

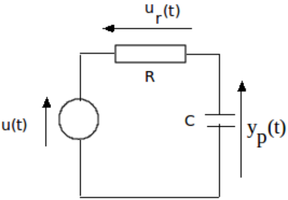
- Système : points de contour
- Modèle : équation du cercle
- Paramètres : centre et rayon

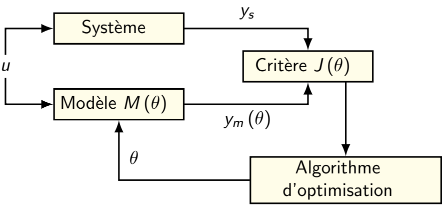






Introduction





Exemple 2

- le processus : le circuit, (entrée $= u(t)$)
- le modèle : l'équation différentielle $y + RCdy/dt = u$
- les observations processus : $y_p(t_n) =$ relevé aux instants t_n
- la sortie modèle $y(t_n) = y_m(t_n) =$ expression mathématique de la solution de l'équation différentielle aux mêmes instants t_n
- le paramètre θ à identifier : $\theta = RC$
- le critère : l'écart entre y_p et y_m,

Introduction

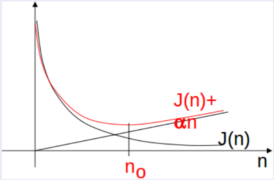
Propriétés des modèles

- Propriétés mathématiques : linéarité (par rapport aux paramètres et/ou aux entrées), dérivabilité...
- Identifiabilité : unicité du vecteur de paramètres pour un comportement donné du modèle.
- Complexité
  - Dimension du vecteur de paramètres
    - + de calculs
    - + de données
    - Modélise le bruit

Linéarité

- $y(k) = \theta u(k)$
- $y(k) = \sqrt{\theta} u(k)$
- $y(k) = \theta u(k)^2$
- $y(k) = \sqrt{\theta} u(k)^2$

Complexité

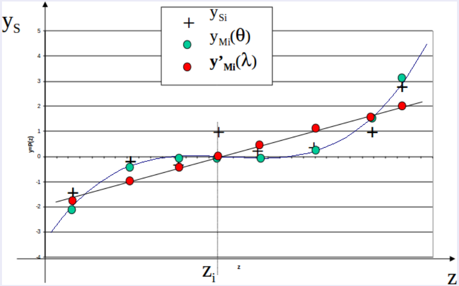


- Akaike $AIC = 2n - ln(L)$

Introduction

Exemple de choix de complexité

- Modélisation polynomiale



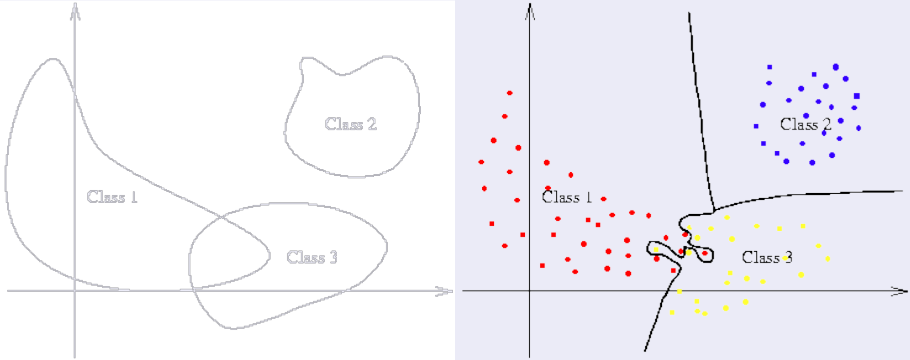
- Exemple Matlab : comment choisir l'ordre K d'un polynôme pour modéliser les données :

$$Y_M(z) = \theta_0 + \theta_1 z + \ldots \theta_K z^K = x^T\theta \quad \text{avec} \quad x^T = [1, z, \ldots z^K]$$

Introduction

Impact de la complexité des modèles en RDF

- Sur-apprentissage

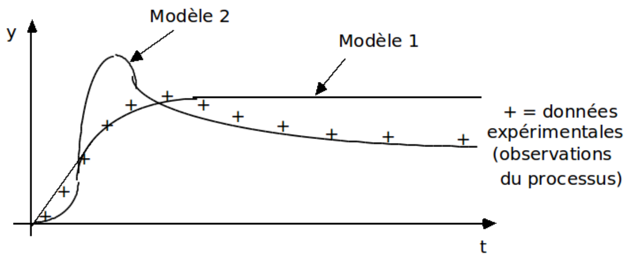


- Problème omniprésent en apprentissage, au cœur des problématique recherche (capacité de généralisation, régularisation, sparsité)

Introduction

Choix des critères

- Doit traduire la ressemblance entre comportement du système et comportement du modèle
- Un modèle $M(\theta_1)$ sera meilleur qu'un autre $M(\theta_2)$ au sens d'un critère J si $J(\theta_1) < J(\theta_2)$
- Notion subjective



- Changer de critère → change de paramètres optimaux

Introduction

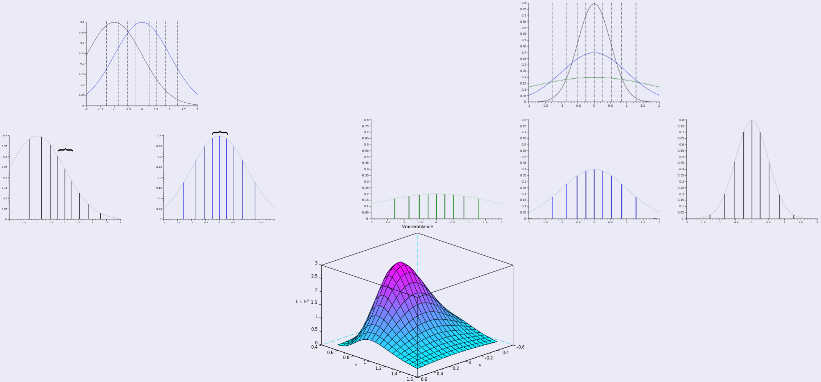
Exemples de critères fréquemment rencontrés

- Critère des moindres carrés : $J_{MC}(\theta) = \sum_n [y_{S\,n} - y_{M\,n}(\theta)]^2$
  - on peut ajouter un facteur de normalisation $1/N$: permet de comparer des valeurs de critères pour des nombre d'observations variables.
  - on peut pondérer les mesures par des w_n positifs. Ces pondérations changent le critère, et donc le minimum. Exemple : permanent/transitoire
- Critère de valeur absolue (norme L^1) ou maximum (norme L^∞)
- Critère du maximum de vraisemblance $J_{MV} = p(y_S|\theta)$: on maximise la "vraissemblance" qu'un jeu de paramètres associés à une distribution ait généré des observations. On considère souvent la log vraissemblance pour des raisons calculatoires.
- Critère du maximum *a posteriori* $J_{MP} = p(\theta|y_S) = \frac{p(y_S|\theta)p(\theta)}{p(y_S)}$
- Les critères peuvent aussi être vectoriels : optimisation multi-objectif

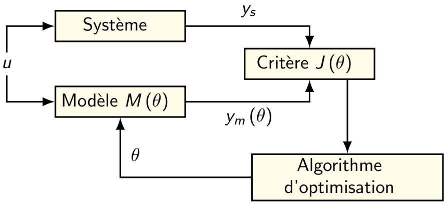
Introduction

Cas du maximum de vraissemblance

- Exemple tiré de Wikipedia : on effectue 9 mesures expérimentales. On les suppose indépendantes et issues d'une loi normale
- La vraissemblance est le produit des probabilités d'apparition de la valeur



Introduction : Bilan sur l'identification de paramètres



Nombreux choix

- Observations
- Modèles
- Critères
- Algorithmes

Et encore bien d'autres

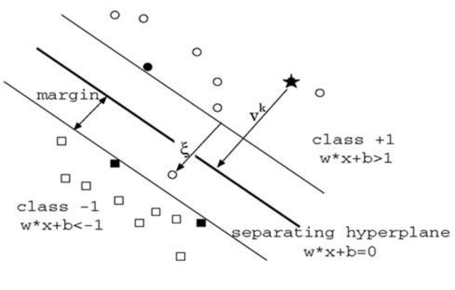
Problèmes numériques, contraintes, modèles dynamiques ...

Introduction

Au delà de l'identification de paramètres

- De nombreux problèmes de RDF impliquent l'optimisation de critères

SVM



- Un critère à minimiser : $J = \frac{1}{2}\left\|w\right\|^2 + C\sum \xi_k$
- Des contraintes : $l_k(w^T x_k + w_0) \geq 1 - \xi_k$

Introduction

Au delà de l'identification de paramètres

- De nombreux problèmes de RDF impliquent l'optimisation de critères

Distances d'édition entre graphes

An attributed graph G is a 4-tuple $G = (V, E, \mu, \xi)$, where :

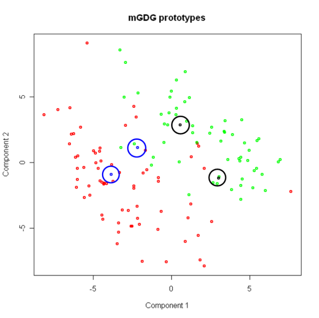
- V is a set of vertices,
- E is a set of edges, such that $\forall e = (i, j) \in E, i \in V \text{ and } j \in V$,
- $\mu : V \rightarrow L_V$ is a vertex labeling function
- $\xi : E \rightarrow L_E$ is an edge labeling function

The graph edit distance $d(.,.)$ is a function

$$
\begin{aligned}
d \quad &: \quad \mathcal{G} \times \mathcal{G} \rightarrow \mathbb{R}^+ \\
&(G_1, G_2) \mapsto d(G_1, G_2) = \min_{o=(o_1,\ldots,o_k)\in\Gamma(G_1,G_2)} \sum_{i=1}^{k} c(o_i)
\end{aligned}
$$

Introduction

Classification de graphes



$$\begin{aligned} mGDG &= \{gdg_{11}, \dots, gdg_{1m}, \dots, gdg_{N1}, \dots, gdg_{Nm}\} \\ &= \underset{\{g_{ik}\}_{i=1,k=1}^{N,m} \subset U}{\arg\min} \Delta\left(T, \{g_{ik}\}_{i=1,k=1}^{N,m}\right) \end{aligned} \tag{1}$$

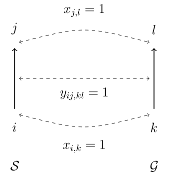
Introduction

Recherche de sous graphes : un problème d'optimisation

- Programmation linéaire en nombres entiers (PLNE)
  - Formulation en « programme mathématique »
  - Résolution par un solveur

Des variables entières

$$\min_{x} \; c^t x$$
$$\text{sous la contrainte } Ax \leq b$$
$$x \in C \subseteq \mathbb{Z}^n$$



Un objectif linéaire

$$\min_{x,y} \left(\sum_{i \in V_{\mathcal{S}}} \sum_{k \in V_{\mathcal{G}}} c_V(i,k) * x_{i,k} + \sum_{ij \in E_{\mathcal{S}}} \sum_{kl \in E_{\mathcal{G}}} c_E(ij,kl) * y_{ij,kl} \right)$$

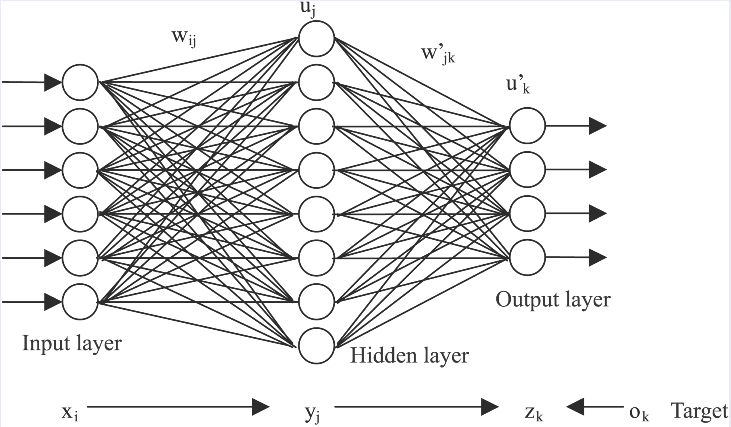
Des contraintes linéaires

$$\sum_{k \in V_{\mathcal{G}}} x_{i,k} = 1 \quad \forall i \in V_{\mathcal{S}}$$
$$\sum_{kl \in E_{\mathcal{G}}} y_{ij,kl} = 1 \quad \forall ij \in E_{\mathcal{S}}$$
$$\sum_{i \in V_{\mathcal{S}}} x_{i,k} \leq 1 \quad \forall k \in V_{\mathcal{G}}$$
$$\sum_{kl \in E_{\mathcal{G}}} y_{ij,kl} = x_{i,k} \quad \forall k \in V_{\mathcal{G}}, \forall ij \in E_{\mathcal{S}}$$
$$\sum_{kl \in E_{\mathcal{G}}} y_{ij,kl} = x_{j,l} \quad \forall l \in V_{\mathcal{G}}, \forall ij \in E_{\mathcal{S}}$$

Introduction

Au delà de l'identification de paramètres

- Exemple des réseaux de neurones

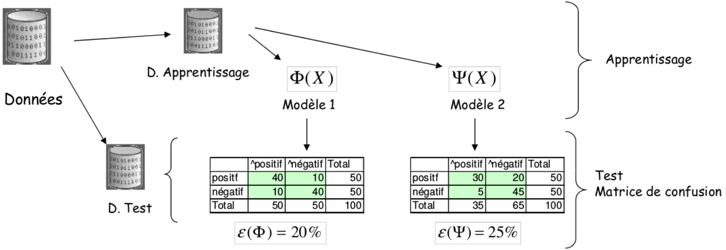


Introduction

Un exemple d'objectifs multiples en apprentissage

- Contexte : problèmes de classification à deux classes
- On a une machine d'apprentissage paramétrée par θ
- On veut choisir la meilleure valeur de θ : sélection de modèles
- Critère classique : taux d'erreur

Schéma habituel d'évaluation des modèles



Introduction

Un exemple d'objectifs multiples en apprentissage

- Problème : si les erreurs n'ont pas la même gravité

Coût de mauvaise affectation non-symétrique

	^positif	^négatif
positf	0	1
négatif	10	0

	^positif	^négatif	Total
positf	40	10	50
négatif	10	40	50
Total	50	50	100

$c(\Phi) = 1.1$

	^positif	^négatif	Total
positf	30	20	50
négatif	5	45	50
Total	35	65	100

$c(\Psi) = 0.7$

Introduction

Un exemple d'objectifs multiples en apprentissage

- Problème identique : si les effectifs sont déséquilibrés
- Exemple du COIL challenge, visant à détecter des personnes qui vont prendre une police d'assurance

0.0650

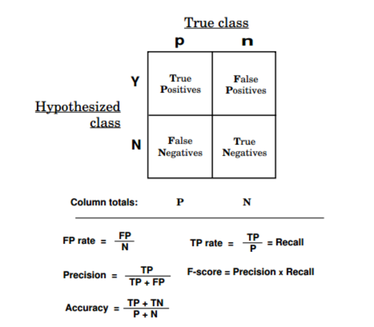
Confusion matrix

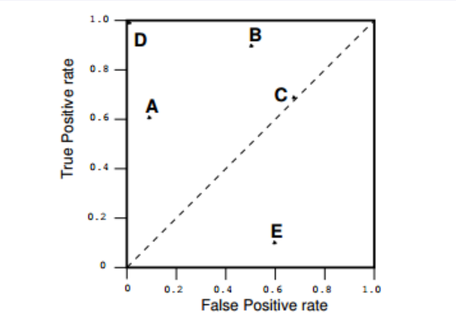
	No	Yes	Sum
No	3731	31	3762
Yes	229	9	238
Sum	3960	40	4000

Introduction

Un exemple d'objectifs multiples en apprentissage : la courbe ROC

- Dans un tel contexte de cost-sensitive classification : il faut distinguer les deux types d'erreur
- L'espace ROC





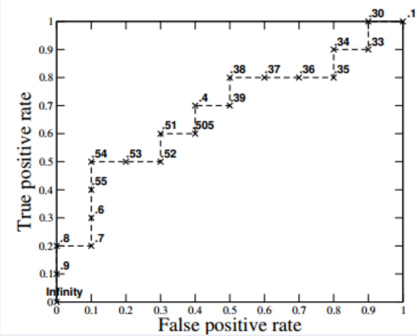
- 2 critères à comparer : Pareto

Introduction

La courbe ROC

- Pour des classifieurs de type mesure : on peut construire une courbe ROC

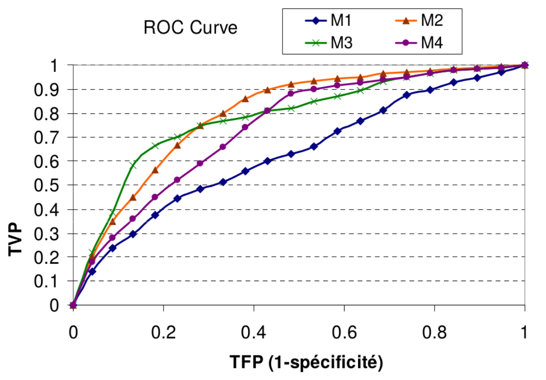
Inst#	Class	Score	Inst#	Class	Score
1	p	.9	11	p	.4
2	p	.8	12	n	.39
3	n	.7	13	p	.38
4	p	.6	14	n	.37
5	p	.55	15	n	.36
6	p	.54	16	n	.35
7	n	.53	17	p	.34
8	n	.52	18	n	.33
9	p	.51	19	p	.30
10	n	.505	20	n	.1



Introduction

La courbe ROC

- Courbe ROC : bon outil de comparaison **cost-sensitive** de classifieurs

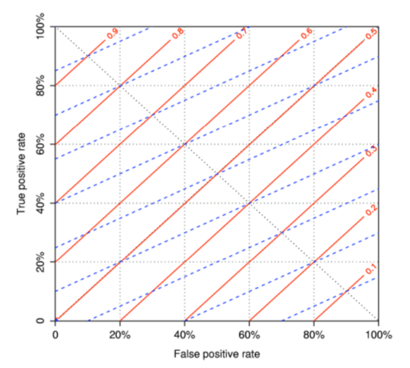


- Critère classique de comparaison : AUC

Introduction

La courbe ROC

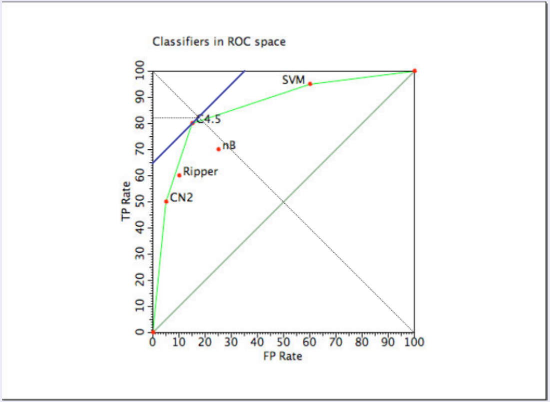
- Fixer le point de fonctionnement : iso-performance line
- Pente d'une iso-performance line :
  - Coûts éq. : $err = pos * (1 - tpr) + neg * fpr \rightarrow p = \frac{neg}{pos}$
  - Coûts deséq. : $cost = pos(1 - tpr)cfn + neg * fpr * cfp \rightarrow p = \frac{neg*cfp}{pos*cfn}$



Introduction

Exemple de choix de classifieur

- $pos = neg$

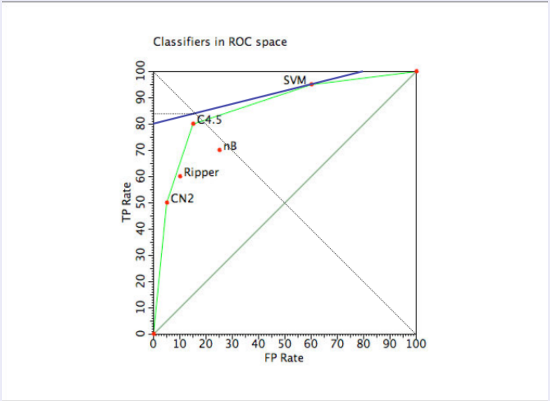


- $acc = 82\%$

Introduction

Exemple de choix de classifieur

- $pos = neg * 4$

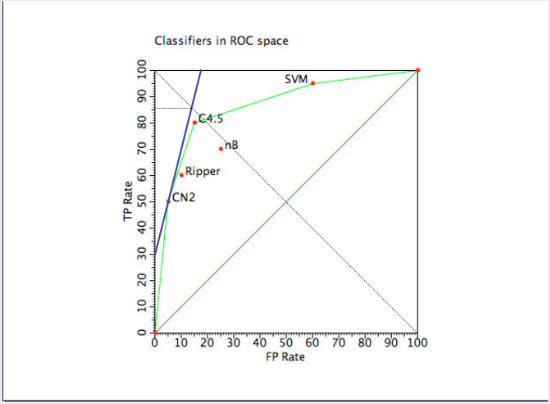


- $acc = 84\%$

Introduction

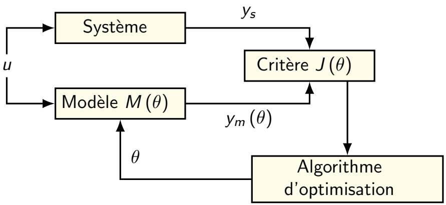
Exemple de choix de classifieur

- $pos = neg/4$



- $acc = 86\%$

Introduction



Plan pour la suite

- Minimum du critère analytiquement calculable : moindres carrés
- Critères continûment dérivables : gradient, Gauss-Newton
- Critères quelconques : recuit simulé, algorithmes évolutionnaires
- Critères multiples
- On relache les contraintes sur le modèle $\rightarrow$ on augmente la complexité

Plan du Cours

Moindres carrés

Hypothèses nécessaires et notations

- Hypothèse 1 : Le modèle Y_M est linéaire par rapport aux paramètres : chaque composante y_{M_i} du vecteur modèle peut s'écrire :

$$y_{M_i} = x_{1i}\theta_1 + x_{2i}\theta_2 + x_{3i}\theta_3 + \ldots + x_{Ki}\theta_K = x_i^T\theta$$

En adoptant une notation matricielle

$$Y_M = \begin{pmatrix} y_{M_1} \\ y_{M_2} \\ \vdots \\ y_{M_N} \end{pmatrix} X = \begin{pmatrix} x_{11} & x_{21} & \ldots & x_{K1} \\ x_{12} & x_{22} & \ldots & x_{K2} \\ x_{13} & x_{23} & \ldots & x_{K3} \\ \ldots & \ldots & \ldots & \ldots \\ x_{1N} & x_{2N} & \ldots & x_{KN} \end{pmatrix} \theta = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_K \end{pmatrix}$$

On écrit :

$$Y_M(\theta) = X\theta$$

Moindres carrés

Vision "algébrique" du problème

- On cherche à résoudre :

$$\left\{\begin{array}{rcl} x_{11}\theta_1 + x_{21}\theta_2 + x_{31}\theta_3 + \ldots + x_{K1}\theta_K & = & y_{S1} \\ x_{12}\theta_1 + x_{22}\theta_2 + x_{32}\theta_3 + \ldots + x_{K2}\theta_K & = & y_{S2} \\ \ldots & & \\ x_{1N}\theta_1 + x_{2N}\theta_2 + x_{3N}\theta_3 + \ldots + x_{KN}\theta_K & = & y_{NN} \end{array}\right.$$

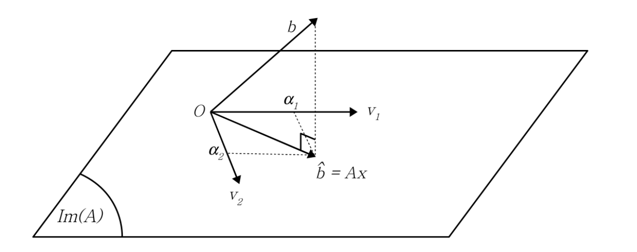
- C'est un système de N équations à K inconnues décrit par $X\theta = Y_S$. Pour $N \leq$K, vous savez résoudre.
- On s'intéresse ici aux systèmes sur-déterminés ($N > K$), pour lesquels il n'y a pas de solution.
- On va chercher à minimiser l'erreur commise

Moindres carrés

Ys与Ym正交

Vision "géométrique" du problème

- On cherche donc à exprimer Y_S (b sur la figure) comme combinaison linéaire des vecteurs colonnes de X (les v_i sur la figure). Cette combinaison appartient à l'espace engendré par ces vecteurs



- On va chercher y_M ($\hat{b}$ sur la figure) tel que $\|Y_S - Y_M\|^2$ ($\left\|\hat{b} - b\right\|^2$) soit minimale : la projection orthogonale de b dans l'espace vectoriel.

Moindres carrés

Vision "analytique" du problème

- On cherche $\theta^T = (\theta_1, \theta_2 ... \theta_K)$ tel que $Y_M = X\theta$ soit le plus proche possible de Y_S
- On voit apparaître la notion de distance (critère).
- Hypothèse 2 : on considère la distance Euclidienne (norme L^2).
- $\theta_{MC} = \arg\min \| Y_S - X\theta \|^2 = \arg\min \sum_{i=1}^{N} \left({y_S}_i - {x_i}^T \theta \right)^2$ avec :

$$\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_K \end{pmatrix} \quad X = \begin{pmatrix} x_1^T \\ x_2^T \\ x_3^T \\ \vdots \\ x_N^T \end{pmatrix} \quad Y_S = \begin{pmatrix} {y_S}_1 \\ {y_S}_2 \\ \vdots \\ {y_S}_N \end{pmatrix}$$

- Remarque : on peut pondérer les différentes observations : $J_{MC}(\theta) = \sum_{i=1}^{N} w_i \left({y_S}_i - {x_i}^T \theta \right)^2 = (Y_S - X\theta)^T W (Y_S - X\theta)$

Moindres carrés

Remarques sur la pondération

- $J_{MC}(\theta) = \sum_{i=1}^{N} w_i \left(y_{S\,i} - x_i^T\theta\right)^2 = (Y_S - X\theta)^T W (Y_S - X\theta)$
- Les w_i pondèrent les erreurs, ils doivent être positifs.
- Si w_i est grand, plus d'importance est accordée à $y_{S\,i}$.
- Exemple 1 : on pondère avec l'indice de la mesure
  - $w_i = t_i$: régime permanent favorisé
  - $w_i = 1/t_i$: régime transitoire favorisé
- Exemple 2 : on pondère par l'écart type σ_i de la mesure i. σ_i caractérise la fiabilité des mesures. Si les erreurs de mesures suivent une loi normale, cette grandeur s'appelle le khi-deux.
  $J_{MC}(\theta) = \sum_{i=1}^{N} \left(\frac{y_{S\,i} - x_i^T\theta}{\sigma_i}\right)^2$
- Dans la suite, pour simplifier les calculs, on posera souvent $w_i = 1 \forall i$ $(W = I)$

Moindres carrés

Résolution du problème de minimisation

- On cherche à minimiser $J_{MC}(\theta) = \|Y_S - X\theta\|^2 = \|X\theta - Y_S\|^2$.
- $J_{MC}(\theta)$ est quadratique $\rightarrow$ le critère est convexe
- On cherche les valeurs de θ telles que $\frac{\partial J_{MC}(\theta)}{\partial \theta} = 0$

Rappels de dérivation

- $\forall x \in \mathbb{R}^k, a \in \mathbb{R}^k, \frac{\partial a^T x}{\partial x} = \frac{\partial x^T a}{\partial x} = a$
- $\forall x \in \mathbb{R}^k, A \in \mathbb{M}_{k,k}, \frac{\partial x^T A x}{\partial x} = (A + A^T)x$

Résolution

- $J_{MC}(\theta) = \|X\theta - Y_S\|^2$
- $\frac{\partial J_{MC}(\theta)}{\partial \theta} =$

Moindres carrés

Résolution du problème de minimisation

- On cherche à minimiser $J_{MC}(\theta) = \|Y_S - X\theta\|^2 = \|X\theta - Y_S\|^2$.
- $J_{MC}(\theta)$ est quadratique $\rightarrow$ le critère est convexe
- On cherche les valeurs de θ telles que $\frac{\partial J_{MC}(\theta)}{\partial \theta} = 0$

Rappels de dérivation

- $\forall x \in \mathbb{R}^k, a \in \mathbb{R}^k, \frac{\partial a^T x}{\partial x} = \frac{\partial x^T a}{\partial x} = a$
- $\forall x \in \mathbb{R}^k, A \in \mathbb{M}_{k,k}, \frac{\partial x^T A x}{\partial x} = (A + A^T)x$

Résolution

- $J_{MC}(\theta) = \|X\theta - Y_S\|^2$
- $\frac{\partial J_{MC}(\theta)}{\partial \theta} = \frac{\partial \theta^T X^T X \theta}{\partial \theta} - \frac{\partial 2\theta^T X^T Y_S}{\partial \theta} + \frac{\partial Y_S^T Y_S}{\partial \theta} = 2X^T X\theta - 2X^T Y_S$
- $\frac{\partial J_{MC}(\theta)}{\partial \theta} = 0 \Rightarrow X^T X\theta = X^T Y_S$

Moindres carrés

Inversion

L'estimée θ_{MC} est donc la valeur de θ solution de l'équation :

$$X^T Y_S = X^T X \theta_{MC}$$

Deux cas de figure :

- $X^T X$ est inversible. On a alors $\theta_{MC} = \left(X^T X\right)^{-1} X^T Y_S$
- $X^T X$ n'est pas inversible. θ_{MC} n'est pas unique.

Cas pratiques de non inversibilité :

- Moins d'observations N que de paramètres K : rang$(X^T X) \leq N < K$ (rang(AB)$\leq$ rang(A) $\leq$ min(dim(A))).
- Dépendance linéaire entre les colonnes de X : on ne peut pas résoudre le système

Dans le cas pondéré, on aura : $\theta_{MC} = \left(X^T W X\right)^{-1} X^T W Y_S$

Moindres carrés (9)

Exemple : Approximation polynomiale

- On choisit de modéliser un processus d'entrée z et de sortie $Y_s(z)$ par un polynôme d'ordre 3 dont on cherche les coefficient optimaux au sens des moindres carrés.
- On relève les mesures suivantes :

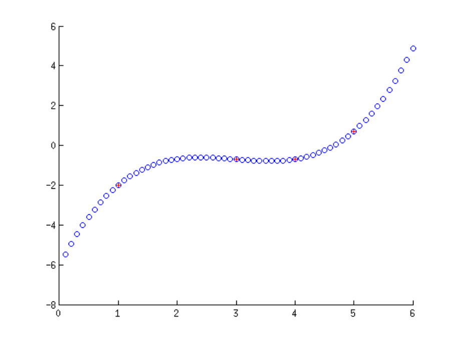
z	1	3	4	5
Y_S	-2	-0.7	-0.7	0.7

- Mettre en équation ce système et donner le calcul à effectuer

Moindres carrés (10)

Exemple : Approximation polynomiale (2)

- On obtient :



- 4 paramètres, 4 observations : on a résolution un système de quatre équation à 4 inconnues.

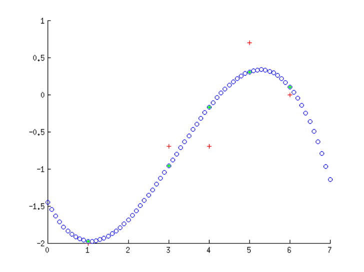
Moindres carrés (11)

Exemple : Approximation polynomiale

- On ajoute une observation :

z	1	3	4	5	6
Y_S	-2	-0.7	-0.7	0.7	0

- 4 paramètres, 5 observations : il n'y a pas de solution exacte. Celle de coût minimal (0.5125) obtenue par les Moindres Carrés donne :



Moindres carrés (12)

Propriétés de l'estimateur

- Modèle : $Y_M = X\theta$
- Minimum du critère $J_{MC}(\theta)$ atteint pour $\theta_{MC} = \left(X^T X\right)^{-1} X^T Y_S$
- On a alors $(Y_M)_{MC} = X\left(X^T X\right)^{-1} X^T Y_S = QY_S$
- L'erreur d'estimation vaut : $Y_S - Y_{M\,MC} = (1 - Q)Y_S$
- Démontrons quelques propriétés de Q :
  - Q est symétrique
  - $Q^2 = Q$
  - $\left(Y_S - QY_S\right)^T QY_S = 0$: $Y_S - (Y_M)_{MC}$ est orthogonal à $(Y_M)_{MC}$
- Retour sur l'interprétation géométrique

Moindres carrés (13)

Interprétation stochastique

- On suppose :
  - $Y_S = Y_M(\theta_S) + b = X\theta_S + b$ où θ_S représente la vraie valeur (inconnue) des paramètres du système.
  - b est un bruit blanc (Gaussien, de moyenne nulle et de variance σ^2) indépendant des x_i
- On appelle erreur d'estimation le vecteur $e_\theta = \theta_{MC} - \theta_S$
- On a $e_\theta = \left(X^T X\right)^{-1} X^T \left(Y_S - X\theta_S\right) = \left(X^T X\right)^{-1} X^T b$
- Donc : $E[e_\theta] = \left(X^T X\right)^{-1} X^T E[b] = 0$
- Et $\mathrm{var}(e_\theta) = E[e_\theta e_\theta^T] = \left(X^T X\right)^{-1} X^T E[bb^T] X \left(X^T X\right)^{-1} = \sigma^2 \left(X^T X\right)^{-1}$

Moindres carrés (14)

Bilan moindre carrés

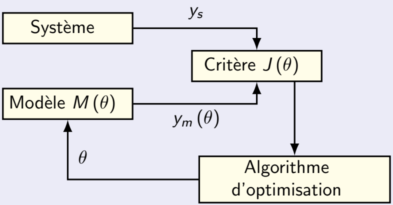
- Modèle linéaire par rapport aux paramètres : $Y_M(\theta) = X\theta$
- Critère erreur quadratique : $J_{MC}(\theta) = \sum_{i=1}^{N} w_i \left(y_{S_i} - y_{m_i}\right)^2$
- Si $X^T X$ est inversible : $\theta_{MC} = \left(X^T X\right)^{-1} X^T Y_S$
- Remarque : en pratique $\left(X^T X\right)$ sera calculé par des méthodes QR
- Estimateur non biaisé le moins dispersé (Gauss-Markov) si la contrainte de linéarité est respectée.
- Modèle généralisable :
  - Si les θ apparaissent dans des expressions telles que $\cos(\theta_1)$, θ_1^2 ... : on change de paramètres.
  - Pour certains cas où Y_M n'est pas linéaire par rapport aux paramètres. Exemple : $Y_M = \theta_1^{\alpha_1}\theta_2^{\alpha_2} \rightarrow ln(Y_M) = \alpha_1 ln(\theta_1) + \alpha_2 ln(\theta_2)$: on change de paramètres et de critère.
- Comment mettre à jour une estimation ?

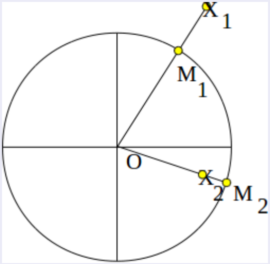
Moindres carrés (15)

Exerçons nous ...

...sur l'exemple 1 de l'introduction

- Observations Y_S = points de contours X_n
- Sorties du modèle $Y_M(\theta)$: ensemble de points M_n





Exemple 1

- $O(a, b)$
- $|O, M_N| = r$
- Paramètres : a, b, r
- Critère : $J(a, b, r) = \sum_n (MnX_n)^2$

Plan du Cours

Plan du Cours

Plan du Cours

Plan du Cours