

Cours de Mathématiques

- ❑ PROBABILITÉS ET STATISTIQUES DESCRIPTIVES
- ❑ MODÈLES LINÉAIRES ET NON LINÉAIRES
- ❑ ALGORITHMES STOCHASTIQUES

Jérôme Lacaille
Expert Émérite Safran

PLAN DU COURS

1. Introduction : réseaux de neurones, applications

- Présentation du plan du cours.
- Statistiques et intelligence artificielle, analyse de données.
- Etudes industrielles.

2. Introduction aux probabilités et statistiques

- Variable aléatoire, échantillon, loi des grands nombres, TCL.
- Estimation, tests, vraisemblance, Neyman-Pearson, maximum de vraisemblance.
- Applications.

Contribution Marie Cottrell, Professeur émérite
université Paris 1 Panthéon-Sorbonne

3. Modèles linéaires

- Régression
- Qualité des estimateurs.
- Sélection de variables.
- Classification.

Contribution Mathilde Mougéot, Maître de
conférences université Paris 7 Diderot.



PLAN DU COURS

4. Modèles non linéaires

- Les arbres de décisions, bagging, boosting et random forest
- Réseaux de neurones, Wiener-Hopf
- Perceptron, rétro-propagation, réseaux récurrents

5. Théorie de l'apprentissage

- Estimation, biais et variance
- Risque empirique, généralisation
- Grandes déviations
- Robustesse

6. Machines de Boltzmann

- Champs de Gibbs, champs markoviens, dynamique de Glauber
- Energie : théorèmes de Dobrushin, Sullivan et Hamersley-Cliford
- Règle de Hebb
- Recuits simulés, algorithmes génétiques
- Réseaux binaires symétriques, calcul parallèle



SUR MOODLE

The screenshot shows the Moodle course index page for 'Ingenieurs MACS an3'. The breadcrumb trail is 'Mes cours > Institut Galilee > Ingenieurs > Ingenieurs MACS an3'. The 'Catégories de cours:' dropdown is set to 'Institut Galilee / Ingenieurs / Ingenieurs MACS an3'. A search bar with 'Rechercher un cours' and a 'Valider' button is present. A list of courses is displayed, with 'Statistiques Descriptives' highlighted by a red dashed box and a red arrow pointing to the URL below.

Accueil Tableau de bord Événements Mes cours

> Mes cours > Institut Galilee > Ingenieurs > Ingenieurs MACS an3

Catégories de cours:

Institut Galilee / Ingenieurs / Ingenieurs MACS an3

Rechercher un cours Valider

Statistiques Descriptives

Images et ondelettes - MACS3

Traitement probabiliste - MACS3

Traitement déterministe - MACS3

<https://moodlelms.univ-paris13.fr/course/index.php?categoryid=197>

The screenshot shows a list of courses. 'Statistiques Descriptives - MACS3' is highlighted with a red box and a red arrow. A detailed view of this course is shown in a pop-up window, listing the course content.

Atelier logiciels - MACS3

Risque de crédit - MACS3

Ingénierie Financière (Mutualise HEC) - MACS3

Projet Bases de données - MACS3

Projet Calcul Parallèle - MACS3

Etude de cas avec OpenTurns - MACS3

EDP non linéaires et Volumes Finis - MACS3 copie 1

HPC et traitement statistique - MACS3

Modélisation des incertitudes - MACS3

Statistiques Descriptives - MACS3


Bases de Données - MACS3




Cours de statistique descriptive pour l'analyse de données industrielles.

1. Introduction : réseaux de neurones, applications
 1. Présentation du plan du cours.
 2. Statistiques et intelligence artificielle, analyse de données.
 3. Etudes industrielles.
2. Introduction aux probabilités et statistiques
 1. Variable aléatoire, échantillon, loi des grands nombres, TCL.
 2. Estimation, tests, vraisemblance, Neyman-Pearson, maximum de vraisemblance.
 3. Applications.
3. Modèles linéaires
 1. Régression
 2. Qualité des estimateurs.
 3. Sélection de variables.
 4. Classification.
4. Modèles non linéaires
 1. Les arbres de décisions, bagging, boosting et random forest



LES ÉLÉMENTS PÉDAGOGIQUES

 > Mes cours > Institut Galilée > Ingénieurs > Ingénieurs MACS an3 > Statistiques Descriptives - MACS3

 |  Annonces 

Modifier ▾

Modifier ▾ 

+ Ajouter une activité ou ressource

 Réseaux de neurones, applications 

Modifier ▾

Une présentation générale d'une méthode d'industrialisation et de cas d'application des réseaux de neurones.

- Statistiques et intelligence artificielle
- L'analyse de données industrielles
- Etudes industrielles

L'idée est de vous montrer de vraies applications industrielles en analyse de données. Le but est de commencer à vous faire réfléchir sur la collaboration avec des ingénieurs ayant des métiers différents du votre mais indispensable pour vous permettre de réaliser des outils qui aideront les bureaux d'études.

  CM 169.1 

Modifier ▾ ☒

+ Ajouter une activité ou ressource

 Probabilités et statistiques 

Modifier ▾



LES COURS SONT TOUJOURS EN LIGNE !

Cours du jeudi 24 septembre à 8h30

SCORE ET INFORMATION DE FISHER

Le score de l'échantillon est défini par

$$s_n(\theta) = \frac{\partial}{\partial \theta} \log L_\theta(X_1, \dots, X_n)$$

- Pour une série d'observations données, le score est une fonction du paramètre θ qui s'annule au maximum de vraisemblance $\hat{\theta}$.
- L'information de Fisher est la variance du score.

$$I_n(\theta) = E(s_n(\theta)^2)$$

- L'information de Fisher mesure l'information apportée par un échantillon sur le paramètre. Si $I_n(\theta)$ est faible, c'est que l'échantillon n'est pas très informatif.
- Le score mesure la sensibilité de la vraisemblance en θ . C'est la dérivée de la vraisemblance donc un score faible montre que la vraisemblance est peu sensible aux variations du paramètre.
- L'information de Fisher peut aussi s'expliquer comme une courbure de la géométrie autour de θ :

$$I(\theta) = -E \left[\frac{\partial^2}{\partial \theta^2} \log L_\theta(X) \right] = -E \left[\frac{\partial}{\partial \theta} s_n(\theta) \right]$$

Intégration par partie de $\int s(\theta)s'(\theta)dP$

- En dimension plus grande que 1, on crée une matrice de Fisher qui représente bien la courbure de géométrie de l'information autour de la valeur du paramètre.

29 / STATISTIQUES DESCRIPTIVES / (2) PROBABILITES ET STATISTIQUES – SAFRAN AIRCRAFT ENGINES



Recorded with BigBlueButton.



Anais Babio:alaoui

Gael Fonguieng Foka:laila Barouk

Anais Babio:damou

Gael Fonguieng Foka:Elmano noel

David Anthony Ouedraogo:ouedraogo david anthony

Gael Fonguieng Foka:laila baroukh

Gael Fonguieng Foka:non elle a deja acces laila

Laila Baroukh:Bonjour

Oumaima Alaoui Mhammedi:Bonjour !

Ismael lyanda Badarou:monsieur, je ne comprends pas quand on dit que le réseau de neurone , c'est juste une descente de gradient

Yanis Zatout:ça a coupé?

Yanis Zatout:Ah non c'est juste ma connexion

Anais Babio:Badarou si j'ai bonne mémoire on l'utilise pour calculer les erreurs de chacun des neurones du réseau

Ismael lyanda Badarou:d'accord, merci anais

Ismael lyanda Badarou:oui monsieur merci



L'EXAMEN DE L'AN DERNIER

Les codes sont disponibles sur GitHub :

<https://github.com/jee51/tabata>

Examen de statistiques descriptives
(Jérôme Lacaille et Florent Forest)

Vous soumettez votre réponse sous la forme d'un fichier notebook (MACS3-SD20-Prenom_Nom.ipynb) que vous enverrez par courrier électronique avant le 1er décembre aux deux adresses suivantes :

- jerome.lacaille@gmail.com
- forest@lipn-univ-paris13.fr.

Ce notebook utilise la toolbox 'tabata' que vous trouverez sur GitHub comme précisé dans le cours : <https://github.com/jee51/tabata>.

L'examen se compose de plusieurs questions, chaque question compte pour un nombre de points défini. La note totale sera la somme des points obtenus par votre présence lors des cours et TD (1/2 point par séance, soit 6 points en tout en comptant les TD), plus les points accumulés par cet examen (20 points), la note finale sera majorée à 20.

Attention cependant, toute journée de retard compte pour un demi point de moins et une recopie évidente (plagiat) du travail d'un de vos camarades compte pour -4 points par contrevenant.

Deadline avant pénalité : le 1er décembre à 0h00 (30 novembre minuit).

Bon travail !

Corrections

Le tableau ci-dessous servira d'évaluation, vous voyez ainsi comment la notation sera découpée en éléments de base. Surtout n'hésitez pas à commenter vos codes et ne laissez pas d'affichages sans légende.

Question	Élément	Points	Note
A Théorie	A.2 Proba/Stats	+1	
	A.3 Modèles linéaires	+1	
	A.4 Modèles non linéaires	+1	
	A.5 Apprentissage	+1	
	A.6 Machines de Boltzmann	+1	
	-----	-----	-----
B Pratique	B.1 Une interface data	+3	
	B.2 Comprendre un algo	+6	
	B.3 Créer un modèle	+6	
-----	-----	-----	-----

Les zones en jaune, rédigées en rouge, ci-dessous sont les questions qui vous sont posées.

A. Questions de cours

Cette partie contient quelques questions de cours permettant de m'assurer que les étudiants qui n'ont pas pu suivre un cours ont visionné



Introduction à Python pour la Data Science

Cours de statistiques descriptives — TD

FLORENT FOREST

✉ forest@lipn.univ-paris13.fr

🌐 <http://florentfo.rest>

🐙 FlorentF9

2020



PRÉSENTATION DES TD

Partie pratique du cours de Statistiques Descriptives (MACS 3).

Objectifs

- > Se familiariser avec le langage Python et les notebooks Jupyter
- > Maîtriser les bases des modules de calcul numérique, d'analyse de données, de visualisation et d'apprentissage (`numpy`, `pandas`, `matplotlib`, `scikit-learn`)
- > Mettre en pratique des méthodes statistiques en réalisant des cas d'études concrets issus du domaine aéronautique



ORGANISATION

Les outils



Aucune installation requise. Seul pré-requis : connexion internet et **compte Google** (pour synchro Drive).

- ▶ github.com/FlorentF9/SupGalilee-tdstats
- ▶ colab.research.google.com
- ▶ framadate.org/SupGalilee-tdstats



PROCÉDURE



► colab.research.google.com

1. Cliquez sur l'onglet GitHub
2. Dans la barre de recherche, tapez FlorentF9, puis sélectionnez le dépôt SupGalilee-tdstats.
3. Cliquez sur un TD pour l'ouvrir (TDX-eleve.ipynb)
4. Vérifiez que vous êtes connecté à votre compte Google (icône en haut à droite).
5. Cliquez sur Copier sur Drive afin de pouvoir sauvegarder vos modifications.
6. Renommer le notebook en ajoutant votre nom (TDX-prenom-NOM.ipynb)
7. Partager avec florent.forest9@gmail.com pour



PRÉSENTATION – JÉRÔME LACAILLE

→ Carrière

1985 - Normalien
1988 - Agrégé de mathématiques
1992 - Docteur en mathématiques : Machines de Boltzmann
1993 - Maître de Conférences ENS Cachan
2004 - Habilité à diriger des recherches : Intégration d'algorithmes

1992 - Fondation de Miriad Parallel Processing
1999 - Directeur R&D de Miriad Technologies
2005 - Directeur Recherche Si-Automation
2006 - Directeur Innovation PDF/Solutions
2008 - Expert Snecma
2010 - Expert Emérite Safran
2015 - Fondation du DATALAB SAFRAN
2016 - Vice président SMAI (Industrie)

→ Enseignements

- Cours de monitorat Licence Orsay (P11).
- TD de vision DEA imagerie.
- Cours Réseaux de neurones (Macs3 / P13)
- Cours de théorie de la mesure.
- Préparation à l'agrégation de statistiques.
- Cours "Réseaux de Neurones" du DEA MVA (Mathématiques Vision Apprentissage).

- Analyse de données (Macs3)



A SUIVRE

INTRODUCTION : RÉSEAUX DE NEURONES ET APPLICATIONS