


XU KAI



ENSAE 3ème année
Stage de fin d'études
Année scolaire 2016-2017

Prévision du défaut sur le périmètre de TPE



Maître de stage : Smaïl IBBOU
du 15 mai au 14 novembre

Table des matières

1	Presentation du groupe BPCE	1
2	Introduction	2
2.1	Contexte	2
2.2	Présentation du sujet	2
3	Traitement des données	4
3.1	Alimentation de la base des données	4
3.2	Traitement des variables	4
4	Apprentissage semi-supervisé	5
4.1	Carte Kohonen (SOM)	5
4.2	Détection de structure topologique	8
4.2.1	Enveloppe convexe	8
4.2.2	Algorithme t-SNE	9
4.3	Carte Kohonen supervisée (SSOM)	10
4.3.1	Sous échantillonnage	10
4.3.2	Modèle de carte Kohonen supervisée	10
4.4	Évaluation des performances	13
4.4.1	Indice de performance	13
4.4.2	Comparaison des performances	14

Liste des tableaux

4.1	Base d'apprentissage	14
4.2	Base de test	14

Table des figures

4.1	Carte Kohonen	5
4.2	Base d'apprentissage	7
4.3	Base de test	7
4.4	Test de séparabilité linéaire	9
4.5	Base de Iris	10
4.6	Base actuelle	10
4.7	Carte Kohonen supervisé	11
4.8	Base d'apprentissage	12
4.9	Base de test	12
4.10	Matrice de confusion : Base de test	13

Chapitre 1

Presentation du groupe BPCE

Le Groupe BPCE s'appuie sur ses deux principaux réseaux, Banque Populaire et Caisse d'Epargne, ainsi que sur ses filiales pour mener à bien toutes les métiers de la banque et de l'assurance. Il est au service de 32 millions de clients avec ses 108 000 collaborateurs. Grâce à ses filiales, le groupe offre à ses clients un service complet : solutions d'épargne et d'investissement, service de placement, de trésorerie, de financement, d'assurance et gestion d'actifs.


Il est le deuxième groupe bancaire en France issu de la fusion en 2009 qui est mise en œuvre en réponse à la crise des subprimes. Les deux réseaux coopératifs gardent leur enseignes et leur indépendance mais coordonne et met en commun leur politique commune. En ce cas, des services de back-office s'intègrent tels que la direction des risques qui comprend l'unité Validation du pôle le Analyses consolidées et Modèles dans laquelle j'ai effectué mon stage de fin d'études.

Chapitre 2

Introduction

2.1 Contexte

Les TPEs (très petites entreprises) constituent un pilier important de l'économie française en occupant environ 20% de l'emploi dans le secteur concurrentiel, et 20% de la valeur ajoutée créée par l'ensemble des entreprises. Actuellement, elles représentent un contingent important du Corporate. Cependant, les modèles actuels **en** prévision de défaut sur la classe d'actifs Corporate ne peuvent pas bien capter ses comportements spécifiques **d'origine de l'effet taille**.

L'apprentissage statistique a été élaboré pour décrire au mieux un phénomène à partir des données volumineuses. **Pourvu** que la population modélisée soit suffisamment grande **pour être** représentative et stable, l'apprentissage statistique sera un moyen idéal à mettre en place afin de décrire au mieux les caractéristiques du défaut de TPE. 

En effectuant une **refonte** des données des réseaux **RBP, RCE** et Natixis, le groupe BPCE a accumulé des données volumineuses de TPE à la fois comportementales et financières. **Du coup**, les données de TPE du groupe possèdent les caractéristiques de quantité volumineuse et de grande dimension, ce qui **lui fait** l'objet potentiel de l'apprentissage statistique.

2.2 Présentation du sujet

La prévision du défaut des clients est l'un des problèmes fondamentaux au cœur de l'attention des banques. L'approche traditionnel consiste à chercher les variables, **p** principalement comptables, qui différencient au mieux les clients défaillants et les clients sains à partir de l'analyse financière. Cependant, un **gap** de performances est observé sur la population de TPE **sous** les modèles existants.

L'objet de notre travail est donc d'appliquer la technique d'apprentissage statistique **dans** les

TPEs afin d'améliorer la prévision du défaut associé à leur comportement spécifique. En ce cas, une meilleure allocation des crédits sera réalisée par une meilleure identification des risques.


Il est possible que les TPEs susceptibles de passer en défaut se distinguent de ses homologues sains par la structure topologique des caractéristiques. c'est pourquoi des méthodes neuronales de carte Kohonen qui permet de faire la classification en préservant la topologie des individus sont implémentées. Basé sur les résultats obtenus de la carte Kohonen, les algorithmes d'apprentissages sont proposés pour ce type de tâches. Ensuite, une méthode hybride est mise en place en fusionnant nos cartes de Kohonen et les méthodes supervisés afin de construire un cadre assez performant et raffiné. Enfin, les validations croisées sont effectuées sur les modèles modifiés en vue de tester la robustesse de nos modèles. Il s'agira d'un côté d'explorer l'espace des entrées et de l'autre d'en développer un cadre prédictif sur les clients TPEs.




Chapitre 3

Traitement des données

3.1 Alimentation de la base des données


Les TPE correspondent à un sous-ensemble des portefeuilles actuels de la Banque de Détail (Retail) Professionnel et Entreprise (Corporate). La critère d'être TPE est que la chiffre d'affaires compris entre 1,5 et 5M€ ou montant total des engagements accordés plus que 1 M€ objet sur lequel nous travaillons est les comptes à vue avec bilan dans la population TPE.

L'organisation du groupe BPCE rend l'alimentation de la base TPE. La base totale est un échantillon occupant environ 1/6 des sociétés qualifiés dans 4 cohortes de Juin 2012 à Décembre 2013 à pas semestriel. La population atypique (associations, entreprises étrangères, SCI, ...) est supprimée de la base. 

Cette base des 18902 effectifs dont environ 3% de clients sont en défaut est divisée en 2 parties : une base d'apprentissage comportant 70% des effectifs qui sert à construire et estimer les modèles différents; une base de test comportant 30% des effectifs utilisée pour tester la stabilité des modèles.



3.2 Traitement des variables

Pour tenter de faire la prévision, nous disposons de variables de 4 catégories : compte, donnée externe, ratios financiers et signalétique. Au total, nous disposons de 202 variables explicatives parmi lesquelles 290 variables quantitatives et 12 variables qualitatives. 

Les variables qualitatives sont mises sous forme de tableau disjonctif pour se transformer en nouvelles variables binaires, ce qui produit environ 56 nouvelle variables. Et les variables quantitatives se normalisent pour être centrées réduites. Du coup, nous travaillons sur 246 variables explicatives générées. En outre, les données manquantes d'une variable seront remplacées par la valeur médiane des données réellement observées si besoin.

Chapitre 4

Apprentissage semi-supervisé

4.1 Carte Kohonen (SOM)

La carte Kohonen est un modèle de réseau neurone artificiel qui nous permet d'apprendre la structure topologique des entrées de manière non supervisée. Elle est formée d'une couche de neurones disposé en grille ou ficelle pour regrouper les entrées. Chaque neurone est caractérisée par un vecteur poids de même dimension que les entrées qui détermine sa position dans l'espace d'entrée. L'apprentissage se déroule d'une manière de la compétition parmi les neurones pour la représentation des entrées qui lui sont présentées. Une partition de Voronoi sera implémenté en vue de découper l'espace d'entrée à partir des vecteurs poids appartenant aux neurones.

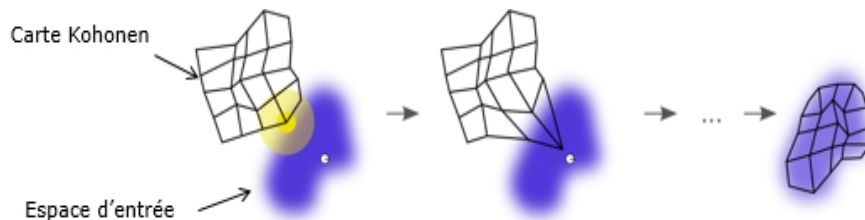


FIGURE 4.1 – Carte Kohonen

Contrairement aux méthodes factorielles telles que l'analyse en composantes principales ou l'analyse de correspondances multiples qui réalisent des projections sur des sous-espaces linéaires de l'espace des entrées, la carte Kohonen capture plutôt la structures topologique des sous-groupes au lieu de l'emplacement précis des individus. D'ailleurs, elle est plus robustes face aux données manquantes [1].

L'algorithme de carte Kohonen [1] :

- 1) Initialisation aléatoire des vecteurs poids \vec{w}
- 2) A instant t , une observation $x(t+1) \in R^d$ est choisie aléatoirement et présentée au réseau
- 3) Le neurone gagnante est définie par :

$$i_0 = \underset{i}{\operatorname{arginf}} \|x(t+1) - w_i(t)\|^2$$

pour la distance euclidienne 4) Les vecteurs poids de neurone gagnante $w_{i_0}^t$ et ses voisins sont mis à jour par :

$$\begin{cases} w_i^{t+1} = w_i^t + \epsilon(t)(x(t+1) - w_i^t) & \forall i \in V_{r_t}(i_0) \\ w_i^{t+1} = w_i^t & \forall i \notin V_{r_t}(i_0) \end{cases}$$

où $V_{r_t}(i_0)$ est le voisinage de rayon $r(t)$ autour de la neurone gagnante i_0 mesuré par la distance de Manhattan $d(\cdot, \cdot)$

$$1_{i \in V_{r_t}(i_0)} = \begin{cases} 1 & \text{if } d(i, i_0) \leq r_t \\ 0 & \text{if } d(i, i_0) > r_t \end{cases}$$

et où $\epsilon(t)$ est le taux d'apprentissage qui satisfait les conditions de Robbins–Monro :

$$\sum_{t=0}^{\infty} \epsilon_t = \infty \quad \text{and} \quad \sum_{t=0}^{\infty} \epsilon_t^2 < \infty$$

Cependant, il se rencontrera deux faiblesses pour cet algorithme [6] : 1) la convergence prématurée ; 2) la mauvaise initialisation. La convergence prématurée est un phénomène dans lequel le réseau de neurones converge si tôt que le résultat se trouve dans un optimum local. Et une mauvaise initialisation pourrait conduire à un réseau moins performant. Pour éviter les deux faiblesses, la quatrième étape du algorithme est modifiée :

$$w_i^{t+1} = w_i^t + \epsilon'(t) e^{\frac{-\|(w_{i_0} - w_i^t)\|^2}{r(t)^2}} (x(t+1) - w_i^t)$$

De plus, une décroissance exponentielle aux hyperparamètres est effectuée en fonction du temps t et du nombre d'itérations T borné :

$$\epsilon'(t) = \epsilon(0) \left(\frac{\epsilon(T)}{\epsilon(0)} \right)^{\frac{T}{t}} \quad r(t) = r(0) \left(\frac{r(T)}{r(0)} \right)^{\frac{T}{t}}$$

Il est évident que la nouvelle mise à jour satisfait aussi la condition de Robbins–Monro :

$$\sum_{t=0}^{\infty} \epsilon'(t) e^{\frac{-\|(w_{i_0} - w_i^t)\|^2}{r(t)^2}} = \sum_{t=0}^T \epsilon'(t) e^{\frac{-\|(w_{i_0} - w_i^t)\|^2}{r(t)^2}} + \sum_{t=T+1}^{\infty} \epsilon(t) = \infty$$

$$\sum_{t=0}^{\infty} \epsilon'(t)^2 e^{\frac{-2\|(w_{i_0} - w_i^t)\|^2}{r(t)^2}} < T\epsilon(0)^2 + \sum_{t=T+1}^{\infty} \epsilon(t)^2 < \infty$$

Après le tuning de paramètre, nous choisissons une 5×5 carte munie des hyperparamètres : $\epsilon(T) = 0.04, \epsilon(0) = 0.01, r(0) = 2.99, r(T) = 0.65$ pour un nombre d'itérations $T=13231$. Nous entendons faire une discrimination en utilisant le regroupement généré par notre carte de Kohonen entraînée. Chaque neurone est marqué d'une classe dont sont muni la plupart des observations représentées par cette neurone.

Le déséquilibre de la répartition du nombre d'observations et de la répartition du défaut est observé dans les résultats de deux bases de données. Conformément à notre attente, la neurone avec des observations le plus nombreuses et ses voisins possèdent les taux de défaut plus bas que ceux des neurones avec des observations moins nombreuses. Cependant, la différence n'est pas si évidente que toutes les neurones sont marquées de la classe saine, ce qui est contre notre attente.

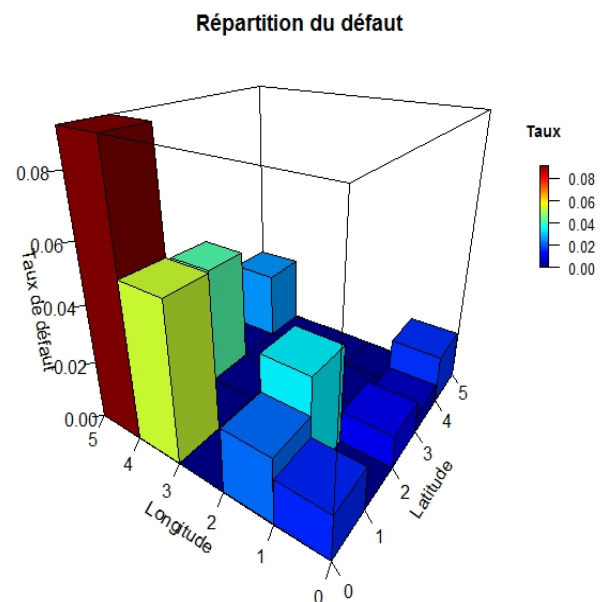
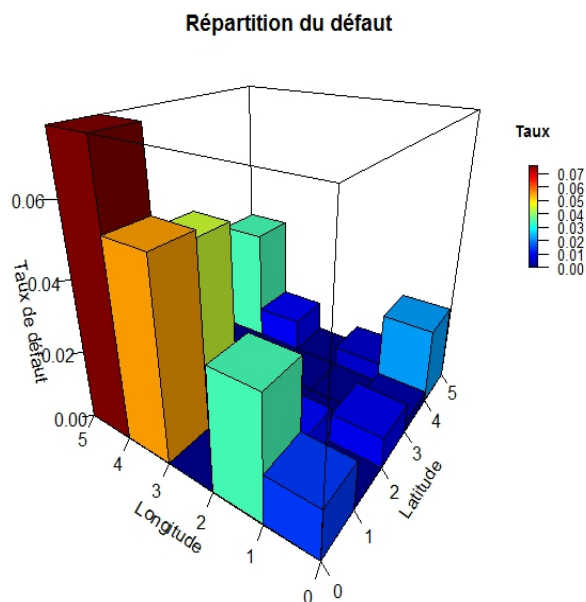
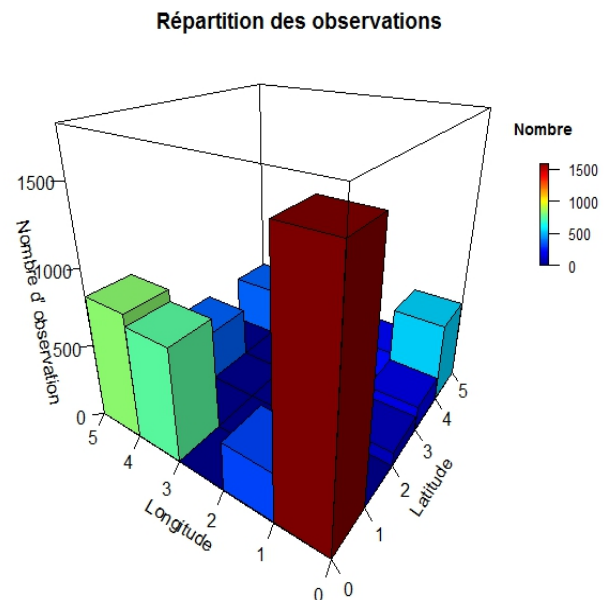
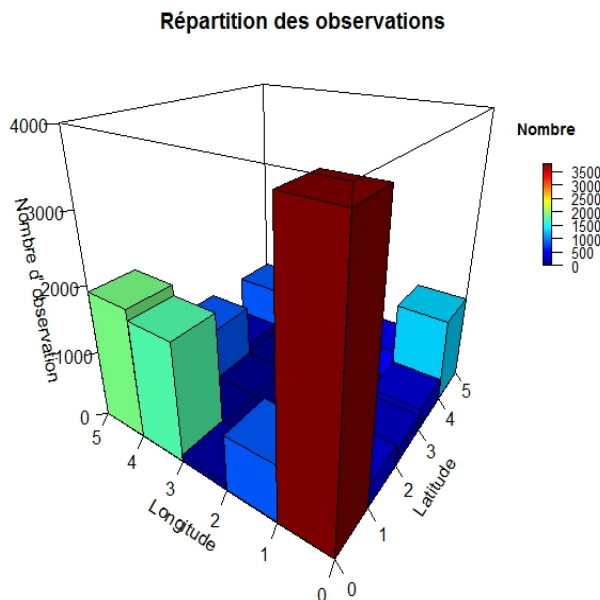


FIGURE 4.2 – Base d'apprentissage

FIGURE 4.3 – Base de test

En fait, ce que nous avons implémenté provient d'une technique de la propagation d'étiquette qui appartient à l'apprentissage semi-supervisé. Selon Chapelle et al [3], la réussite de l'apprentissage semi-supervisé réside dans trois hypothèses : 1) les entrées regroupées dans le même groupe tendent à être de même classe ; 2) l'hyperplan qui sépare les entrées de classes différentes devrait traverser une zone des entrées à faible densité ; 3) les entrées de haute dimension se tombent approximativement sur un objet géométrique de basse dimension.

En ce cas, notre découverte surprenante est possible d'être expliquée par les écarts entre les hypothèses et la structure topologique de nos entrées réelles. Nous explorons donc la structure topologique de nos entrées.

4.2 Détection de structure topologique

4.2.1 Enveloppe convexe

Pour un problème de classification à deux classes, il faut examiner tout d'abord la séparation linéaire pour vérifier la validité des discriminants linéaires. La séparation linéaire est une caractéristique selon laquelle les données de deux classes peuvent être classifiées par un hyperplan. C'est à dire, pour une population \vec{x} de deux classes, il existe un vecteur \vec{w} et un seuil c qui permettent le hyperplan $\vec{w} \cdot \vec{x} = c$ sépare les deux classes.

L'enveloppe convexe d'une classe est l'ensemble convexe le plus petit parmi ceux qui le contiennent. Si les enveloppes convexes se croisent, les deux classes se voir comme inséparables linéairement. Ici, l'analyse en composantes principales est effectuée pour visualiser les enveloppes convexes en 2 dimensions.



La figure 3.1 révèle que les classes ne sont pas séparables linéairement parce que les enveloppes se chevauchent. Cette découverte nous montre que l'analyse discriminante linéaire ne peut pas être appliquée à nos données. De plus, les méthodes plus avancées permettant la séparation non-linéaire seront implémentées.

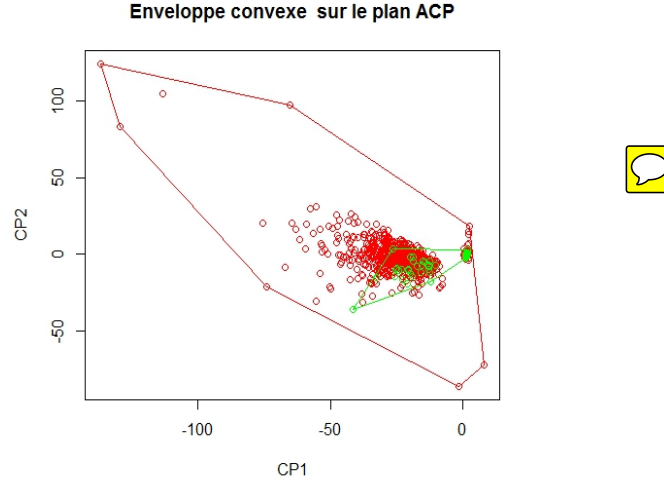


FIGURE 4.4 – Test de séparabilité linéaire

4.2.2 Algorithme t-SNE

Pour démontrer la faiblesse de la carte Kohonen non-supervisée qui ne profite que la structure topologique, une visualisation des données est effectuée par l'algorithme t-SNE (t-distributed stochastic neighbor embedding). Pour les points à grande dimension $\mathbf{x}_1, \dots, \mathbf{x}_N$, une distribution de probabilité p_{ij} est définie à partir de la probabilité conditionnelle $p_{j|i}$ s'agissant de la similarité des paires des individus :

$$p_{j|i} = \frac{\exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-\|\mathbf{x}_i - \mathbf{x}_k\|^2 / 2\sigma_i^2)}$$

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2N}$$

Une distribution de probabilité q_{ij} est également définie de la même manière pour les points projetés $\mathbf{y}_1, \dots, \mathbf{y}_N$:

$$q_{j|i} = \frac{\exp(-\|\mathbf{y}_i - \mathbf{y}_j\|^2)}{\sum_{k \neq i} \exp(-\|\mathbf{y}_i - \mathbf{y}_k\|^2)}$$

$$q_{ij} = \frac{q_{j|i} + q_{i|j}}{2N}$$

L'algorithme t-SNE est destiné à trouver une projection optimale des points à grande dimension dans un espace de 2 dimension en minimisant la divergence de Kullback-Leibler entre les deux distributions. De cette manière, la proximité des individus est bien gardée. Les points projetés $\mathbf{y}_1, \dots, \mathbf{y}_N$ sont déterminés de manière suivante :

$$\inf_{\mathbf{y}_1, \dots, \mathbf{y}_N} KL(P||Q) = \sum_{i \neq j} p_{ij} \log \frac{p_{ij}}{q_{ij}} \quad (4.1)$$

Le succès de la classification en utilisant la structure topologique réside dans la séparation topologique des individus de classes différentes. Beaucoup de bases de données de l'apprentissage machine

profite de cette propriété, par exemple, la Base de Iris. Cependant la population de notre base ne manifeste la séparation provenant de la structure topologique, ce qui devient une faiblesse de la carte Kohonen non-supervisée. Les figures suivantes illustrent cette faiblesse.

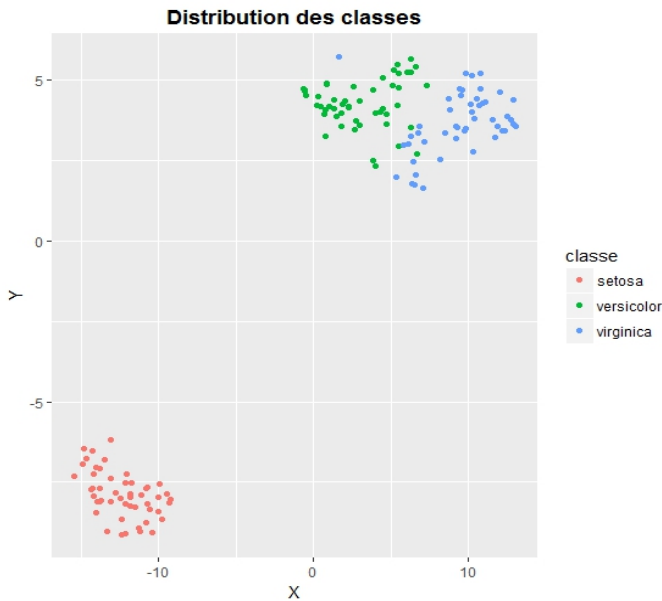


FIGURE 4.5 – Base de Iris

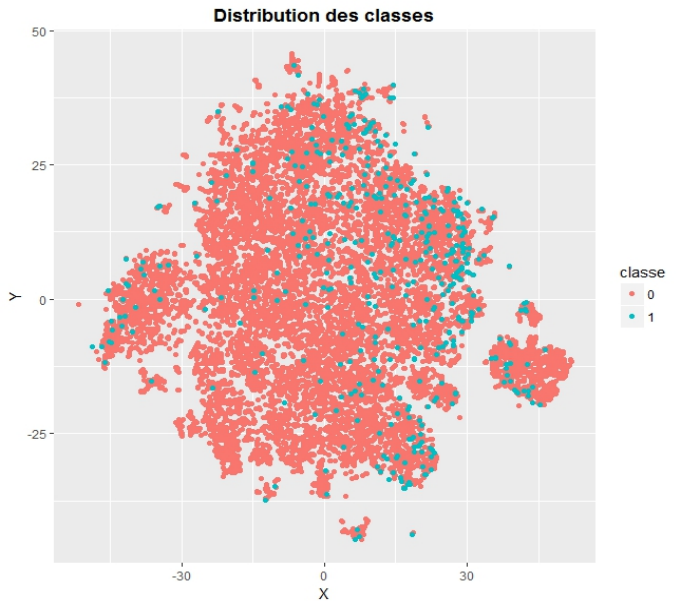


FIGURE 4.6 – Base actuelle



4.3 Carte Kohonen supervisée (SSOM)

La section précédente nous montre que seulement la structure topologique n'est pas suffisant pour faire la classification. Le déséquilibre des classes nous pose aussi un problème. Notre regroupement nécessite les informations supplémentaires et les traitements particuliers. C'est pourquoi le sous échantillonnage et la carte Kohonen supervisée sont effectués.

4.3.1 Sous échantillonnage

Le sous échantillonnage est une technique qui lutte contre le déséquilibre des classes en diminuant des observations de classe majoritaire dans la base d'apprentissage.[5] Il est préférable d'utiliser cette méthode lorsque la base de données est énorme et la réduction du nombre des entrées aide à améliorer le temps d'exécution et les problèmes de stockage. Ici, un sous-échantillonnage des observations de classe majoritaire est réalisé par tirage aléatoire selon une loi uniforme.



4.3.2 Modèle de carte Kohonen supervisée



D'après Melssen et al [6], la carte Kohonen supervisée est un réseau fusionné qui profite à la fois des entrées et des classes à l'étape d'apprentissage. Ce réseau se décompose en deux types de couches : couches d'entrée et couches de sortie. Ici, nous pouvons décomposer de plus les couches

d'entrée en deux nouveaux types : couches d'entrée quantitative et couches d'entrée qualitative. De cette manière, on obtiendra trois types de couches au lieu de deux types de couches en modèle traditionnel. Le réseau s'entraîne de même manière que celui non supervisé sauf que une pondération des couches est possible.

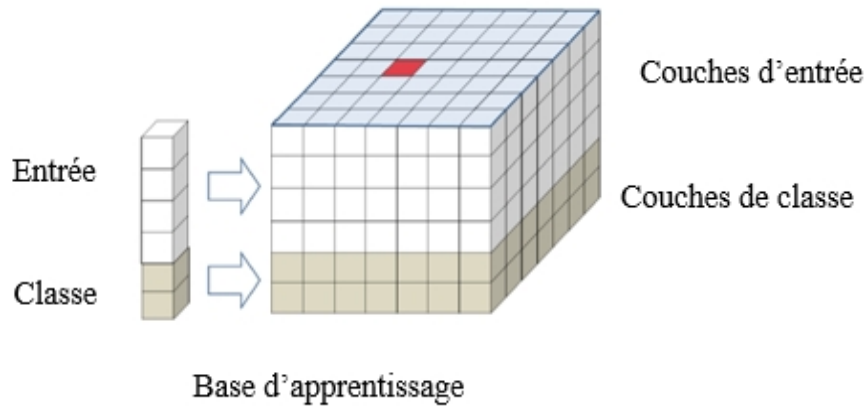


FIGURE 4.7 – Carte Kohonen supervisé

Nous effectuons tout d'abord une sous échantillonnage en diminuant le nombre des clients sains dans la base d' apprentissage à environ 1.1 fois celui de clients défaillants. Ensuite, un **tuning** de paramètre est réalisé. Pour la simplicité, nous supposons les poids parmi les couches de même type sont pareils. De ce fait, on ne fait que le **tuning** de paramètre sur les poids de différent type. Enfin, un réseau dont un poids de $1/2$ est affecté aux couches d'entrée et un poids de $1/2$ est affecté aux couches de sortie se trouve le plus raisonnable. En plus, le poids affecté aux couches d'entrée se décompose en deux parties : $3/8$ pour les couches d'entrée quantitative et $1/8$ pour celles d'entrée qualitative.

La carte Kohonen non supervisée ne peut pas produire les neurone marqué de classe défaillante. Par contre, la carte Kohonen supervisée combinée avec le sous échantillonnage est capable de produire des neurones défaillantes, ce qui est démontré dans les plafonds des figures de boîte suivantes. Le reste des observations sont classifiées selon la classe de neurone qui les représente. Le résultat de la classification sur des deux bases complètes se démontre dans les figures suivantes. Cette fois, la relation inverse entre le nombre des observations de neurone et le taux de défaut local est plus évident. Le taux de défaut augmente sur les neurones défaillantes malgré un manque de la dominance du côté de nombre des clients défaillants.

D'ailleurs, une matrice de confusion est construit sur la base de test pour manifester la qualité de classification de la carte Kohonen supervisée. Chaque colonne de la matrice représente le nombre d'occurrences d'une classe réelle lorsque chaque ligne représente le nombre d'occurrences d'une classe estimée.

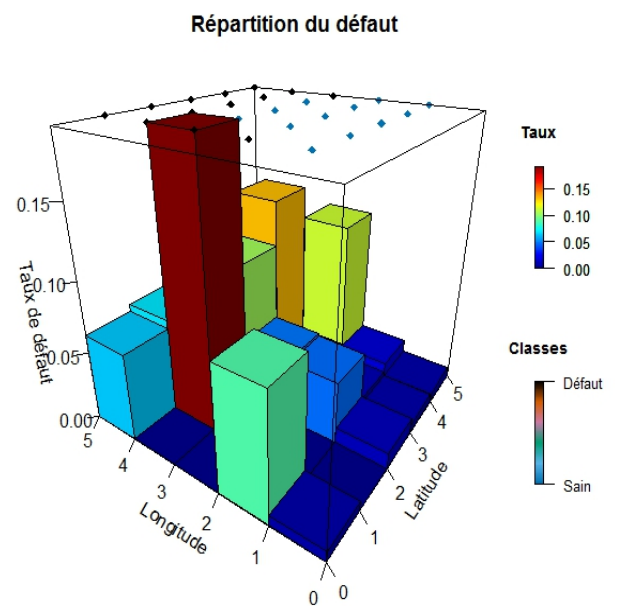
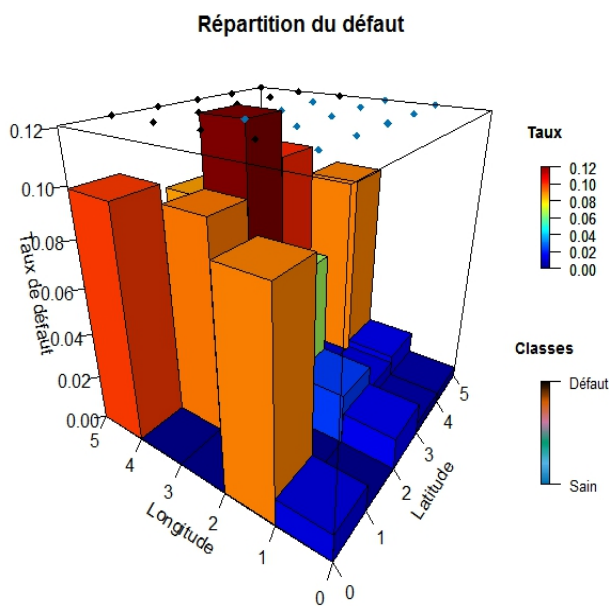
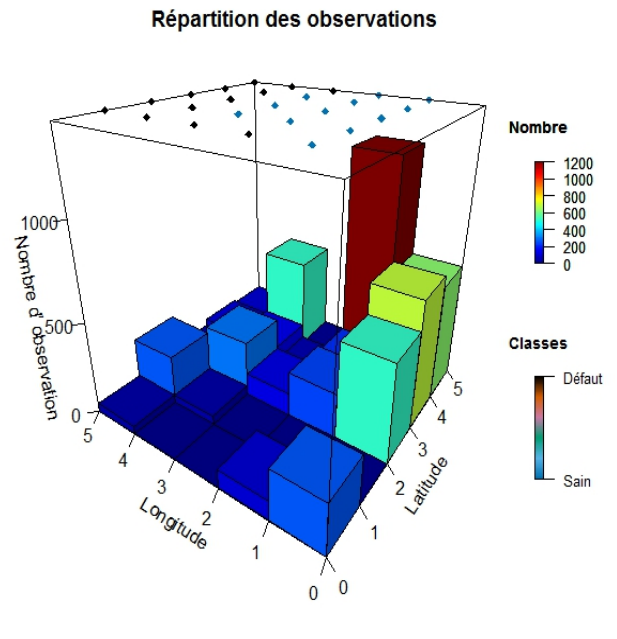
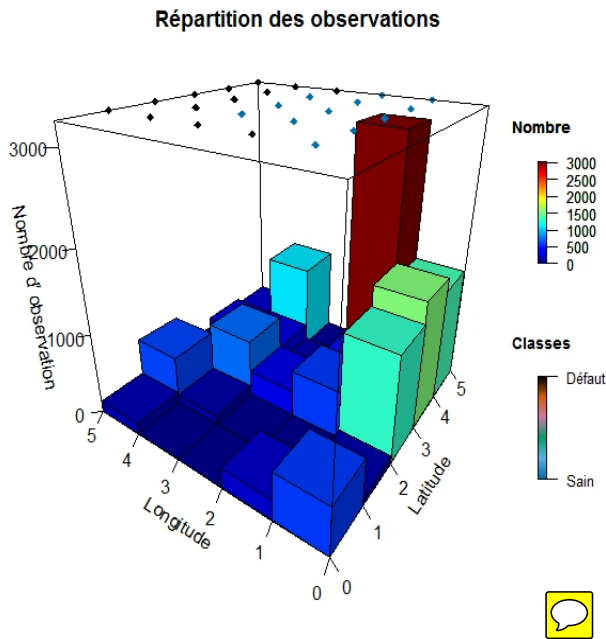


FIGURE 4.8 – Base d'apprentissage

FIGURE 4.9 – Base de test

La matrice nous donne les informations suivantes : sur les 5480 clients sains, 3887 seront estimés comme sains, soit 71% des clients sains sont correctement prédits ; et sur les 189 clients défaillants, 133 seront estimés comme défaillants, soit 70.4% des clients sains sont correctement prédits. Ce résultat est de bonne performance, mais il nous reste à voir s'il existe des modèles traiter mieux des informations fournies.



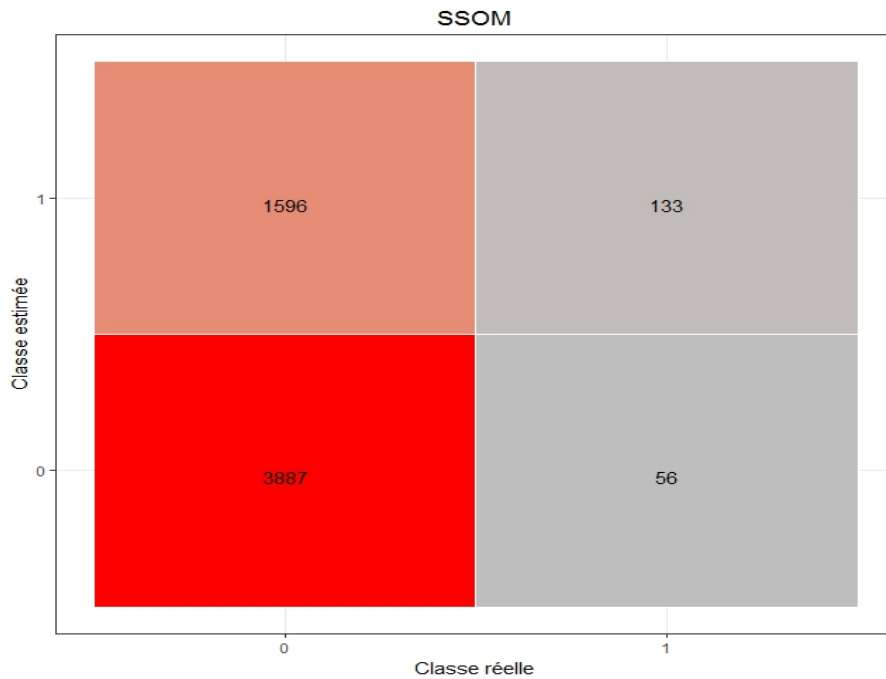


FIGURE 4.10 – Matrice de confusion : Base de test

4.4 Évaluation des performances



4.4.1 Indice de performance

Au cours du regroupement, il nous faut évaluer la qualité de regroupement pour la comparaison et le tuning de paramètre. Les méthodes d'évaluation doivent être compétentes de mesurer la performance des réseaux générés et tolérer le cas où les neurones marqués de classe défaillante manquent comme on l'a vu avec l'échec de la carte Kohonen non supervisée. Du coup, trois critères sont choisis pour faire l'évaluation : pureté, information mutuelle normalisée et . Sauf l'explication spécifique, les partitions ci-dessous sont deux à deux disjoints par défaut.



Pureté

La pureté est un critère qui évalue la mesure dans laquelle les groupes contiennent une classe unique. Compte tenu d'un ensemble de groupes M et d'un ensemble de classes D , tous les deux divisent N individus, tous les deux divisent N individus. La pureté se calcule de manière suivante :

$$\frac{1}{N} \sum_{m \in M} \max_{d \in D} |m \cap d|$$



Information mutuelle normalisée (NMI)

Contrairement à la pureté qui préfère le grand nombre de groupes, l'information mutuelle normalisée pénalise le grand nombre de groupes. En ce cas, la taille de la carte Kohonen se contrôle. Elle mesure la similarité entre deux partitions d'un ensemble. Compte tenu d'un ensemble de groupes U de R groupes et d'un ensemble de classes V de C classes, les entropies des deux partitions se définissent par :

$$H(U) = - \sum_{i=1}^R P(i) \log P(i); H(V) = - \sum_{j=1}^C P'(j) \log P'(j)$$

Et l'information mutuelle de deux partitions se définit par :

$$MI(U, V) = \sum_{i=1}^R \sum_{j=1}^C P(i, j) \log \frac{P(i, j)}{P(i)P'(j)}$$

$$NMI(U, V) = \frac{MI(U; V)}{[H(X) + H(Y)]/2}$$

Indice de Rand (RI)

L'indice de Rand mesure aussi la similarité entre deux partitions d'un ensemble. L'enjeu de ce critère est de mesurer la consistance (le taux d'accord) entre deux partitions. Nous notons deux partitions U et V :

$U \setminus V$	Co-groupée	Non co-groupée
Co-groupée	a	b
Non co-groupée	c	d

L'indice de Rand se calcule par :

$$RI(\pi_1, \pi_2) = (a + d) / \binom{n}{2}$$

4.4.2 Comparaison des performances

Indice \ Méthode	SOM	SSOM	SSOM*
Pureté	0.967	0.967	1
NMI	0.00179	0.0126	0.421
RI	0	0.00647	0.185

Tableau 4.1 – Base d'apprentissage

Indice \ Méthode	SOM	SSOM
Pureté	0.967	0.967
NMI	0.00283	0.0119
RI	0.00145	0.00647

Tableau 4.2 – Base de test

SSOM* : Carte Kohonen supervisée sur la base d'apprentissage sous échantillonnée

Selon les tableau ci-dessus, l'effet de la combinaison du sous échantillonnage et l'apprentissage supervisé de SSOM est vraiment évident. Néanmoins, il nous faut remettre les individus exclus par le sous échantillonnage pour faire la prévision. En ce cas, l'effet de sous échantillonnage s'affaiblit et une baisse de la performance est observée. Cependant, l'effet du apprentissage supervisé de SSOM reste dans tous les deux bases de données, ce qui permet une meilleure performance de SSOM sur la base complète que celle de SOM classique.

Dans le chapitre suivant, nous explorerons l'apprentissage supervisé pour étudier comment profiter mieux de l'interaction entre des entrées et les classes. Dans le chapitre 5, nous introduirons une méthode hybride qui essaie de garder l'effet de sous échantillonnage en utilisant la carte Kohonen lorsque l'apprentissage se déroule de manière supervisée.

Bibliographie

- [1] IBBOU Smail *Classification, analyse des correspondances et methodes neuronales*, . These,Universite Paris 1, 1992.
- [2] Melssen, Willem, Ron Wehrens, and Lutgarde Buydens. *Supervised Kohonen networks for classification problems*. Chemometrics and Intelligent Laboratory Systems 83.2 (2006) : 99-113.
- [3] Chapelle, Olivier, Bernhard Scholkopf, and Alexander Zien. *Semi-Supervised Learning*. MIT Press, 2006
- [4] Yuan-chin Ivan Chang *Boosting SVM classifiers with logistic regression*. www.stat.sinica.edu.tw/library/citec/ep/2003-03.pdf,2003
- [5] Dai, Dong, and Shaowen Hua. *Random Under-Sampling Ensemble Methods for Highly Imbalanced Rare Disease Classification*. Proceedings of the International Conference on Data Mining (DMIN), 2016.
- [6] Melssen, W., Wehrens, R., Buydens, L. *Supervised Kohonen networks for classification problems*. Chemometrics and Intelligent Laboratory Systems, 83(2), 99-113 (2006).
- [7] Chawla, Nitesh V. *Data mining for imbalanced datasets : An overview*. Data mining and knowledge discovery handbook. Springer US, 2009. 875-886.
- [8] Schütze, H. *Introduction to information retrieval* . Proceedings of the international communication of association for computing machinery conference.(2008, June).
- [9] Louppe, G., Wehenkel, L., Sutura, A., Geurts, P. *Understanding variable importances in forests of randomized trees*. . Advances in neural information processing systems (pp. 431-439), (2013).
- [10] Kuhn, M., Johnson, K. . *Applied predictive modeling* New York, NY : Springer (2013)
- [11] Richard G. Baraniuk *Compressive Sensing*. IEEE Signal Processing Magazine
- [12] Xu, Z., Huang, G., Weinberger, K. Q., Zheng, A. X. *Gradient boosted feature selection*. Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 522-531). ACM. (2014, August).
- [13] Meinshausen, N., Bühlmann, P. *Stability selection*. Journal of the Royal Statistical Society : Series B (Statistical Methodology), 72(4), 417-473.(2010).
- [14] Breiman L., Friedman J. H., Olshen R. A., and Stone, C. J. *Classification and Regression Trees*. (1984) Wadsworth.

- [15] Boczko, E. M., Xie, M., Wu, D., and Young, T. *Comparison of binary classification based on signed distance functions with support vector machines*. 2009. OCCBIO'09. Ohio Collaborative Conference on (pp. 139-143). IEEE.
- [16] Xu, Qing-Song, and Yi-Zeng Liang. *Monte Carlo cross validation*. Chemometrics and Intelligent Laboratory Systems 56.1 (2001) : 1-11.