



Mars 2018
Wei GE
Yi QIAO

Optimisation d'un portefeuille d'actifs par les réseaux de neurones

Encadrant :

Jean-Marc LE CAILLEC
Didier GUERLOT



IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

Table des matières

1	Introduction	3
1.1	Problématique	3
1.2	Présentation de données	3
2	Conception et développement de modèle	4
2.1	Construction de la base d'apprentissage	4
2.1.1	Pré-traitement de données	4
2.1.2	Paramétrage de la base d'apprentissage	4
2.2	Choix du modèle	4
2.3	Préparation des scénarios	4
3	Résultats obtenus et analyse	6
3.1	Résultats de différents scénarios	6
3.1.1	Variation de taille d'apprentissage	6
3.1.2	Variation de l'horizon	7
3.2	Evaluation du modèle	7
3.3	Difficultés rencontrées	7
4	Prospectives et points à améliorer	8
5	Conclusion	9

1 Introduction

Dans le marché financier, il est très important de faire des analyses sur la portefeuille d'instrument pour optimiser le PnL global. Cependant, nous avons aussi tenu compte des risques liés aux opérations financières. Dans le cadre de notre projet, nous avons étudié sur la portefeuille d'actifs et nous avons utilisé les données historiques pour prédire le rendement dans un horizon prédéfini.

L'objectif de ce rapport est pour vous présenter notre projet en détail. Il y a 5 parties dans ce rapport en global. D'abord, nous avons expliqué le contexte du projet, nous avons montré notre problématique et les données fournies par nos encadrants. Et puis, nous avons pris le modèle de réseaux neurones pour réaliser la prédiction, donc nous avons discuter sur la construction de la base d'apprentissage et les paramètres du modèle. Nous avons présenté nos résultats par scénario dans la troisième partie et aussi nous avons évalué notre modèle par les indicateurs statistiques. Dans la partie suivante, nous avons donné des points à améliorer pour la continuation de ce projet. A la fin, il y a une conclusion synthétique.

1.1 Problématique

Un portefeuille d'actifs est construit par plein d'actions avec les poids différents, les actions sont impacté par plusieurs facteurs comme le secteur, le pays, etc. Pour gérer un portefeuille d'actifs, l'objectif est de maximiser le rendement global, c'est-à-dire qu'il faut être capable de faire la décision d'acheter ou de vendre des actions pour optimiser le rendement de chaque action. Dans le cadre de notre projet, nous avons besoin d'optimiser d'un portefeuille CAC40. Comme les séries temporelles sont non stationnaires, donc il est très difficile de faire la prédiction pour chaque période.

1.2 Présentation de données

Le portefeuille CAC40 est construit par 49 actions françaises et chaque action a ses comportements complètement différents. Nous avons un fichier qui contient le prix (ouverture, clôture, haut, bas) de chaque action pendant presque 15 ans (2000-2015). Pour une série temporelle, si nous avons pris une période trop courte, nous n'avons pas assez d'informations pour faire la prédiction. Par contre, elle n'a pas une forte corrélation pendant une longue période, les données très anciennes ne sont pas utile de prédire la situation actuelle.

Nous avons aussi un fichiers ayant les 13 « technical indicators » calculés à partir de ces séries temporelles. Les TIs se sont situés dans 4 catégories : le volume, la volatilité, la tendance, le momentum. Dans un premier temps, nous avons pris tous les 13 TIs, et puis nous avons étudié la corrélation entre eux pour diminuer le nombre de TIs.

2 Conception et développement de modèle

Nous avons trouvé qu'il y a déjà beaucoup de projets sur la prédiction d'une séries temporelle financière. Pour atteindre notre objectif, nous avons proposé le modèle de réseaux neurones pour faire la prédiction. L'entrée et la sortie de notre projet sont très différentes par rapport aux autres projets. Les entrées sont les 13 TIs, comme nous savons plus de TIs peuvent apporter plus d'informations, mais il y a aussi plus de risques d'avoir trop de redondances. La sortie est le rendement sur un horizon prédéfini, c'est plus difficile de prédire la vraie valeur que la tendance (mettre un seuil d'achat ou de vente). Dans cette partie, nous avons d'abord présenté la façon de construction de la base d'apprentissage, nous avons fait un pré-traitement sur les données origines et nous avons paramétré la base d'apprentissage.

2.1 Construction de la base d'apprentissage

2.1.1 Pré-traitement de données

Comme nous avons pris un apprentissage supervisé, nous avons besoin d'abord d'ajouter les labels pour nos exemples, c'est-à-dire nous avons calculé le rendement sur un horizon pour chaque exemple. Le formule de calculer le rendement est $r = \frac{P_{t_0+h_0}}{P_{t_0}} - 1$.

Dans un premier temps, nous avons utilisé directement les données non normalisées, parce que nous avons considéré que notre modèle peut ajuster les valeurs par lui-même, c'est-à-dire que la différence produit par l'échelle va être compensé automatiquement par les poids. Cependant, quand nous avons fait le premier test sur les données non normalisé, nous avons trouvé que le résultat n'était pas idéal. Le système n'a donné que un constant négatif et un constant positif comme le résultat, il n'a pas fait un rapprochement vers la vraie valeur. Ensuite, nous avons appliqué une normalisation aux entrées, nous avons pris le formulaire $(V - V_{min}) / (V_{max} - V_{min})$, ça vaut dire que nous avons mis toutes les valeurs de l'entrée sur la fourchette de [0,1]. Après la normalisation, nous pouvons trouver que le résultat était mieux et la valeur de prédiction a approché vers la vraie valeur.

2.1.2 Paramétrage de la base d'apprentissage

Pour notre projet, il faut aussi varier sur les paramètres différents pour construire la base d'apprentissage. Principalement, nous avons pris 4 indicateurs pour mesurer notre système, ils sont l'horizon, la taille d'apprentissage, la durée du test et le nombre de TI.

Pour lancer les simulations différentes, nous avons d'abord changé seulement un indicateur sur les tests différents. Et puis, nous pouvons tester sur le cas où nous pouvons changer 2 ou 3 indicateurs pour trouver le meilleur combinaison.

Nous avons pris tous les 13 TIs au début de notre projet, nous voudrions collecter plus d'information pour faire la prédiction, nous avons considéré que cette combinaison de TI est plus robuste. Pour savoir le performance de notre système, nous avons diminué le nombre de TI, nous avons refait les tests avec la dimension réduite de TI. Pour l'instant, nous avons varié le nombre de TI sur 13, 12 et 4.

2.2 Choix du modèle

Nous avons utilisé l'algorithme de perceptron multicouche à rétropropagation. Le perceptron multicouche est un type réseau organisé par plusieurs couches, chaque couche est constituée d'un nombre variable de neurones, les neurones de la dernière couche étant les sorties du système global. Dans notre cas, l'entrée de réseau est les TIs et la sortie est le rendement. L'algorithme de rétropropagation du gradient est pour minimiser l'erreur dans les systèmes multicouches. le figure ?? pour représenter la structure de notre réseau neurone :

Pour être plus clair, nous avons présenté nos exemples d'entrées en format matriciel :

2.3 Préparation des scénarios

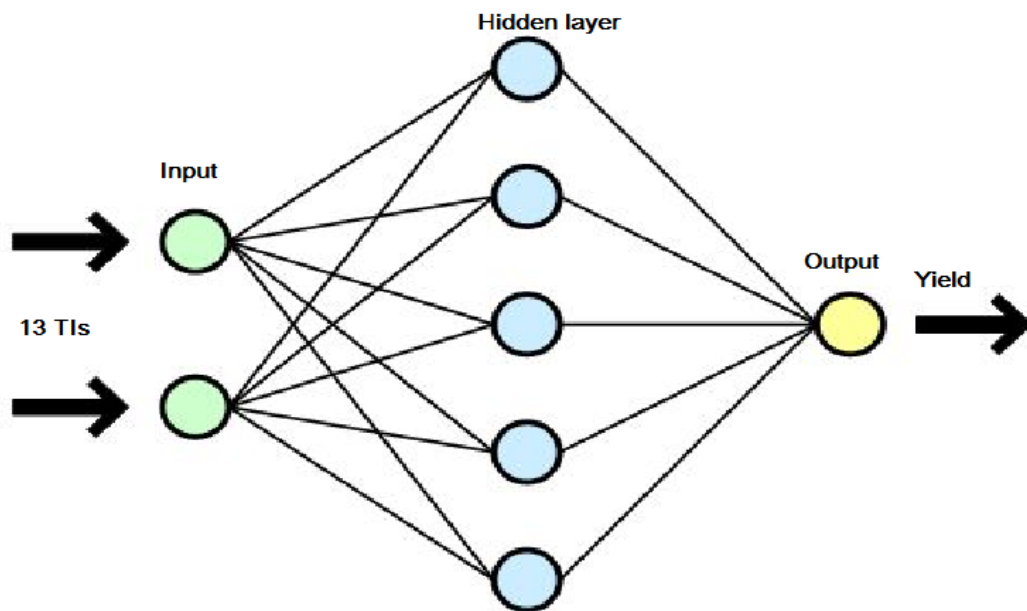


FIGURE 1: Le schema de réseaux neurones

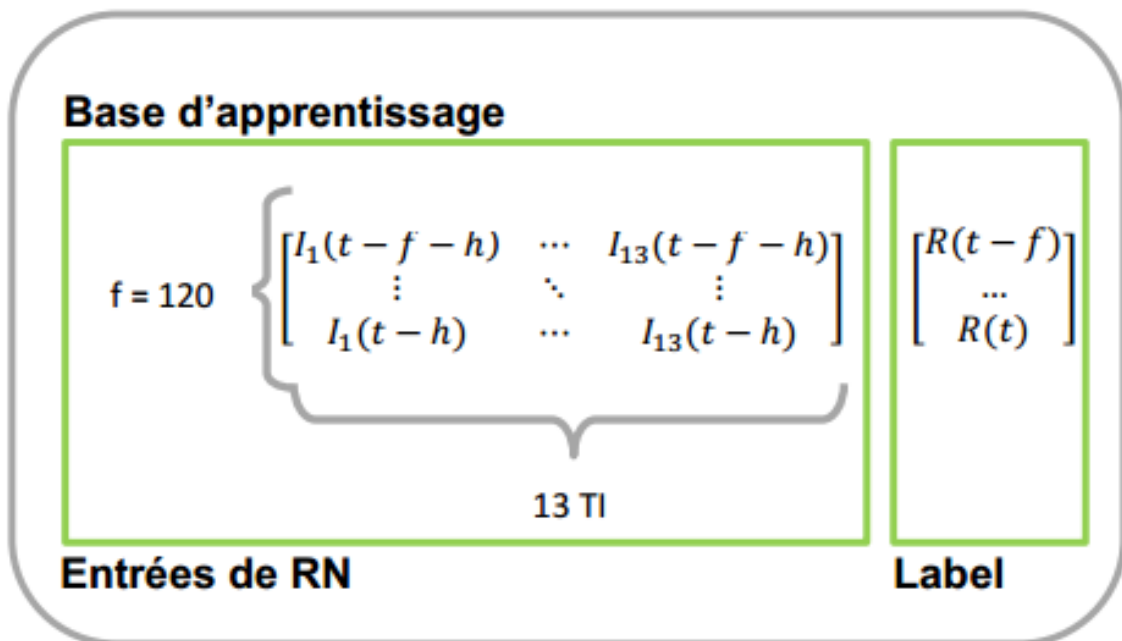


FIGURE 2: L'entrée de base d'apprentissage

3 Résultats obtenus et analyse

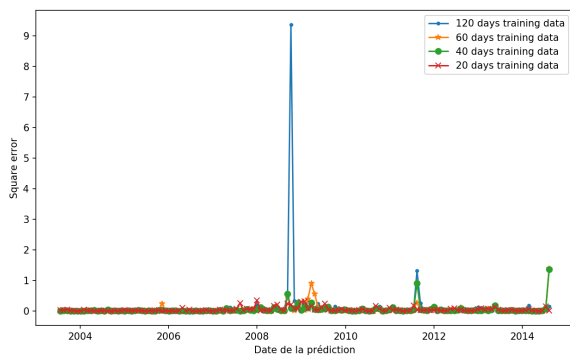
3.1 Résultats de différents scénarios

3.1.1 Variation de taille d'apprentissage

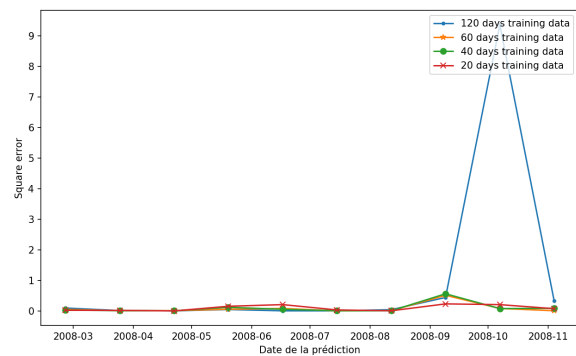
Comme nous travaillons sur des séries temporelles, qui ne sont pas stationnaires, nous nous intéressons à l'effet de la taille d'apprentissage sur la précision de la prédiction du rendement de l'action. Nous avons réalisé des tests sur l'action Renault dans l'indice CAC40.

Sur la figure 3, nous observons un très grand pic au début d'octobre 2008 avec une taille d'apprentissage de 120 jours (6 mois), juste après le début de la crise mi-septembre. Toutefois, ce pic est absent sur les autres courbes. Cela signifie que notre modèle est capable de détecter le début de la période de la crise, puisqu'une fois que celle-ci est arrivée, la tendance du marché est brutalement rompue. Dans ce cas, apprendre sur les données des 6 derniers mois n'est pas une bonne approche pour la prédiction du rendement. En revanche, apprendre sur une période plus courte est plus efficace pour ne prendre en compte que les informations intéressantes de l'évolution du marché.

Dans les autres périodes, nous pouvons constater que l'erreur quadratique de la prédiction du rendement est proche de zéro dans les quatre cas.



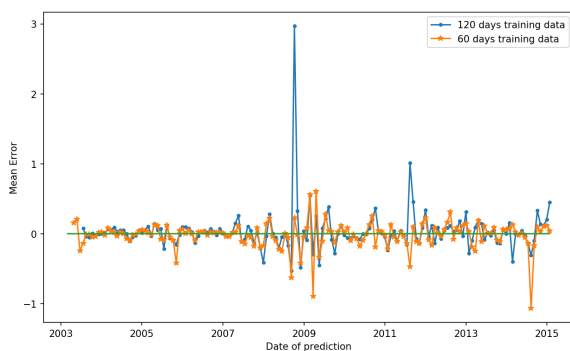
(a) Vue globale sur 10 ans



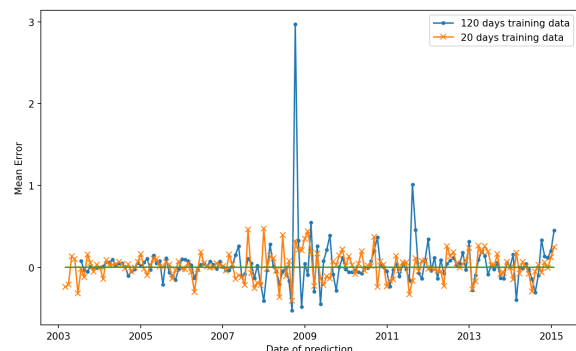
(b) Vue zoomée sur la période de la crise de 2008

FIGURE 3: Erreur quadratique de la prédiction du rendement de Renault à un horizon de 20 jours (1 mois) en fonction de la taille d'apprentissage

En observant les figures ci-dessous, nous remarquons que l'erreur moyenne pour 120 jours est généralement plus petite que les autres cas (sauf la période de la crise), car une période plus longue contient plus d'informations et peut mieux présenter la tendance du marché. Au contraire, une période plus courte est plus bruitée.



(a) Erreur moyenne pour 120 et 60 jours



(b) Erreur moyenne pour 120 et 20 jours

FIGURE 4: Erreur Moyenne de la prédiction du rendement de Renault à un horizon de 20 jours (1 mois) en fonction de la taille d'apprentissage

3.1.2 Variation de l'horizon

3.2 Evaluation du modèle

3.3 Difficultés rencontrées

L'apprentissage a pris trop de temps.

Trop de scénarios

Bcp d'actions

4 Prospectives et points à améliorer

La décision (à seuil, achat, vente, ne rien à faire)
Appliquer à toutes les actions dans le portefeuille

5 Conclusion