



Trading Automatique

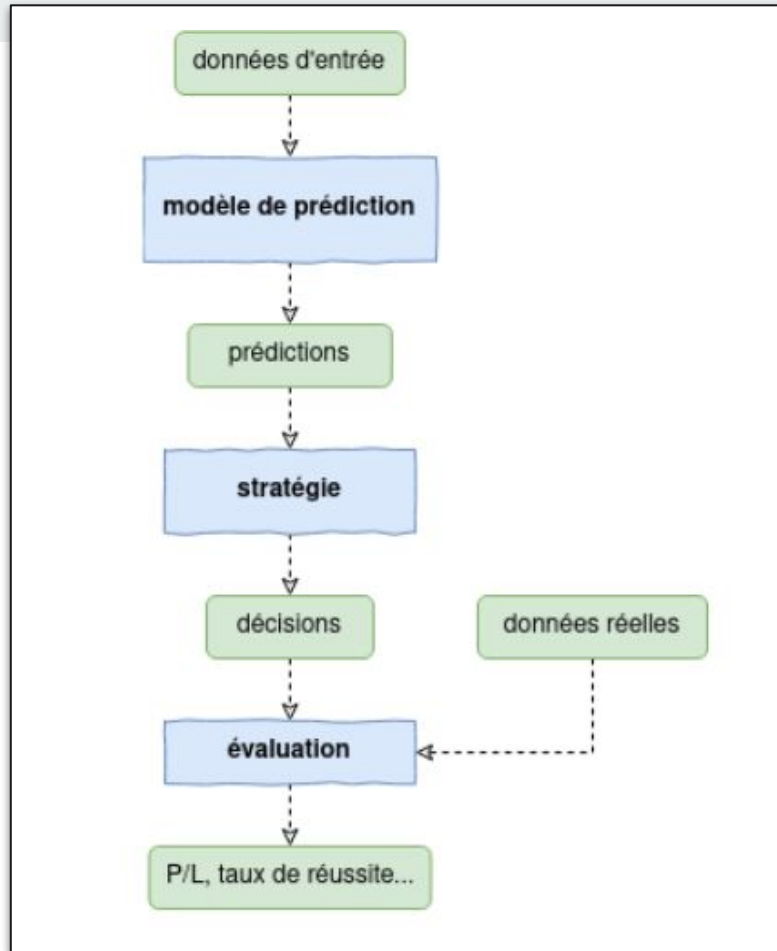
Hugo Kyhm

Matthieu Nogatchewsky



Plan

1. Mise en contexte
 - a. Qu'est ce que le trading automatique?
 - b. Passer de la prédiction à la stratégie
 - c. Limite des métriques actuelles
2. Problématique
 - a. Feuille de route
 - b. Ressources à disposition
 - c. Backtesting
3. Coeur du projet
 - a. Modèles de prédiction
 - b. Approche par classification : réfutation des métriques actuelles
 - c. Approche par classification : réfutation des loss function usuelles
 - d. Métrique et fonction de perte d'exploitabilité réelle
4. Résultats : Comparatif de notre métrique vs. la MSE



Mise en contexte

Le trading automatique c'est optimiser :
le modèle ET la stratégie

$$\forall m \in \mathcal{M} \quad \forall s \in \mathcal{S} \quad e(s^* \circ m^*(x), y) \geq e(s \circ m(x), y)$$

L'évaluation est fixée à des mesures type
P&L, Winning Rate, Equity curve...

Les sources bibliographiques sont peu
répandues, un des problèmes les plus
difficiles en machine learning



Plan

1. Mise en contexte
 - a. Qu'est ce que le trading automatique?
 - b. Passer de la prédiction à la stratégie**
 - c. Limite des métriques actuelles
2. Problématique
 - a. Feuille de route
 - b. Ressources à disposition
 - c. Backtesting
3. Coeur du projet
 - a. Modèles de prédiction
 - b. Approche par classification : réfutation des métriques actuelles
 - c. Approche par classification : réfutation des loss function usuelles
 - d. Métrique et fonction de perte d'exploitabilité réelle
4. Résultats : Comparatif de notre métrique vs. la MSE

De la prédiction à la stratégie



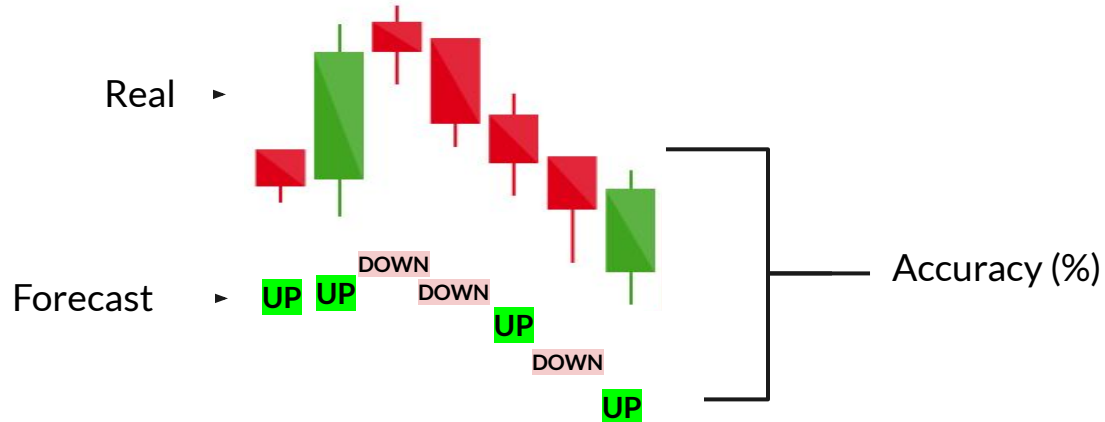


Plan

1. Mise en contexte
 - a. Qu'est ce que le trading automatique?
 - b. Passer de la prédiction à la stratégie
 - c. **Limite des métriques actuelles**
2. Problématique
 - a. Feuille de route
 - b. Ressources à disposition
 - c. Backtesting
3. Coeur du projet
 - a. Modèles de prédiction
 - b. Approche par classification : réfutation des métriques actuelles
 - c. Approche par classification : réfutation des loss function usuelles
 - d. Métrique et fonction de perte d'exploitabilité réelle
4. Résultats : Comparatif de notre métrique vs. la MSE

Limites des métriques actuelles

- Les papiers de recherches communiquent principalement les résultats sur l'accuracy
- Pour les classifieurs, la sortie est binaire: UP ou DOWN



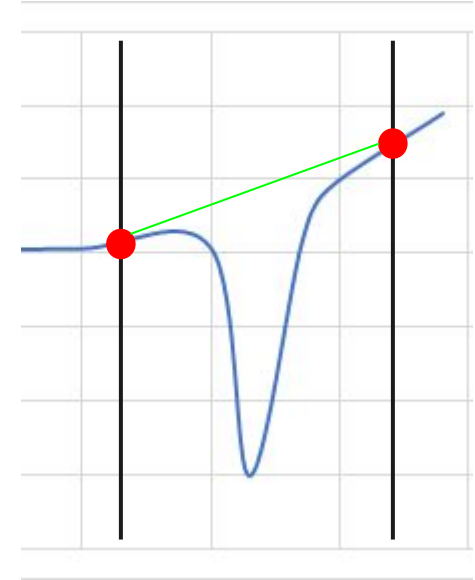
Limites des métriques actuelles

- Certains publient des accuracy de l'ordre de 60% soit un peu plus qu'une prédiction aléatoire
- Or une accuracy haute n'implique pas une stratégie profitable

Quelques problèmes liés à l'accuracy



Magnitude



Risk exposure



Plan

1. Mise en contexte
 - a. Qu'est ce que le trading automatique?
 - b. Passer de la prédiction à la stratégie
 - c. Limite des métriques actuelles
2. **Problématique**
 - a. **Feuille de route**
 - b. Ressources à disposition
 - c. Backtesting
3. Coeur du projet
 - a. Modèles de prédiction
 - b. Approche par classification : réfutation des métriques actuelles
 - c. Approche par classification : réfutation des loss function usuelles
 - d. Métrique et fonction de perte d'exploitabilité réelle
4. Résultats : Comparatif de notre métrique vs. la MSE



Problématique

Mesurer l'exploitabilité réelle dès l'entraînement

1. Modèles de prédiction et limites des métriques usuelles
2. Métrique et fonction de perte rendant de la profitabilité potentielle
3. Application et comparaison de notre/nos métriques avec les métriques usuelles



Plan

1. Mise en contexte
 - a. Qu'est ce que le trading automatique?
 - b. Passer de la prédiction à la stratégie
 - c. Limite des métriques actuelles
2. Problématique
 - a. Feuille de route
 - b. Ressources à disposition**
 - c. Backtesting
3. Coeur du projet
 - a. Modèles de prédiction
 - b. Approche par classification : réfutation des métriques actuelles
 - c. Approche par classification : réfutation des loss function usuelles
 - d. Métrique et fonction de perte d'exploitabilité réelle
4. Résultats : Comparatif de notre métrique vs. la MSE

Ressources à disposition

- Base de données:
 - Historique de prix sous forme OHLCV
 - 57 instruments financiers
 - 8 ans d'historiques 2010 à 2017
 - Intervalles de temps (bar) 5min à 4j
- Outil de backtesting:
 - Stratégie embarquée: à chaque prédiction DOWN => SELL / UP => BUY
 - Permet de tester rétroactivement les performances du modèles
 - P&L



	open	high	low	close	volume
2010-01-03 22:00:00	1.43318	1.43318	1.42969	1.42994	49
2010-01-03 22:05:00	1.42994	1.43014	1.42974	1.42974	22
2010-01-03 22:10:00	1.42974	1.43063	1.42974	1.43063	30
2010-01-03 22:15:00	1.43063	1.43068	1.43058	1.43063	8
2010-01-03 22:20:00	1.43063	1.43158	1.43063	1.43156	53



Plan

1. Mise en contexte
 - a. Qu'est ce que le trading automatique?
 - b. Passer de la prédiction à la stratégie
 - c. Limite des métriques actuelles
2. Problématique
 - a. Feuille de route
 - b. Ressources à disposition
 - c. **Backtesting**
3. Coeur du projet
 - a. Modèles de prédiction
 - b. Approche par classification : réfutation des métriques actuelles
 - c. Approche par classification : réfutation des loss function usuelles
 - d. Métrique et fonction de perte d'exploitabilité réelle
4. Résultats : Comparatif de notre métrique vs. la MSE



Backtesting

$$e_{P\&L} = \sum_i q \times \epsilon(\hat{y}_i) \times y_i = \sum_i q \times \epsilon(y_i \times \hat{y}_i) \times |y_i|$$

	All	Long	Short
Net P&L	5194.00	11736.00	-6542.00
Gross Profit	24912.00	19584.00	5328.00
Gross Loss	-19718.00	-7848.00	-11870.00
Profit Factor	1.26	2.50	0.45
Total # of Trades	84.00	48.00	36.00
Number Winning Trades	35.00	26.00	9.00
Number Losing Trades	49.00	22.00	27.00
Percent Profitable	0.42	0.54	0.25
Avg Trade Win Loss	61.83	244.50	-181.72
Avg Winning Trade	711.77	753.23	592.00
Avg Losing Trade	-402.41	-356.73	-439.63
Ratio Avg Win Loss	1.77	2.11	1.35
Largest Winning Trade	2169.00	2169.00	1723.00
Largest Losing Trade	-1530.00	-764.00	-1530.00
Opened P&L	-620.00	-620.00	0.00



Plan

1. Mise en contexte
 - a. Qu'est ce que le trading automatique?
 - b. Passer de la prédiction à la stratégie
 - c. Limite des métriques actuelles
2. Problématique
 - a. Feuille de route
 - b. Ressources à disposition
 - c. Backtesting
3. **Coeur du projet**
 - a. **Modèles de prédiction**
 - b. Approche par classification : réfutation des métriques actuelles
 - c. Approche par classification : réfutation des loss function usuelles
 - d. Métrique et fonction de perte d'exploitabilité réelle
4. Résultats : Comparatif de notre métrique vs. la MSE



Modèles de prédiction

- Prise en main de la base de données
- Support pour tester nos métriques
- Pas d'objectif de performance
- Corrélation avec le P&L, profitabilité du modèle
- Distinction entre l'approche par classification et par régression



Plan

1. Mise en contexte
 - a. Qu'est ce que le trading automatique?
 - b. Passer de la prédiction à la stratégie
 - c. Limite des métriques actuelles
2. Problématique
 - a. Feuille de route
 - b. Ressources à disposition
 - c. Backtesting
3. Coeur du projet
 - a. Modèles de prédiction
 - b. Approche par classification : réfutation des métriques actuelles**
 - c. Approche par classification : réfutation des loss function usuelles
 - d. Métrique et fonction de perte d'exploitabilité réelle
4. Résultats : Comparatif de notre métrique vs. la MSE



Approche par classification

- Paramètres
 - Instrument EURUSD
 - Subdataset size: 3000 (90% training / 10% test)
 - Bar size: 5 min
 - Lookback: 30
 - Modèles : Bayesian, CNN, Decision Tree, Random Forest
- Résultats
 - Accuracy proche de 50%
 - Profitabilité avec variance élevée



Formalisation d'un protocole d'évaluation des métriques

Comparing the metrics

We compare the metrics by putting them into a table with the following format:

	Combination 1	...	Combination 16	Average	Correlation
Metric 1					
...					
Metric n					
Backtest (P&L)					

We use one row for each metric, and the last row is for the backtest (P&L for example).

For each metric, we put the value of this metric for each combination of parameters (optional), the average of all the combinations and the correlation of the average and the backtest value in the columns.



Corrélation des métriques classiques

Métrique	Corrélation avec le P&L
Accuracy	0.07
MAE	0.06
F1 micro	0.01
Precision micro	0.04
Recall micro	0.001

Points faibles de la classification

- Prédiction binaire UP/DOWN
- Pas de prise en compte de l'amplitude des variations
- Nécessité de considérer les modèles de régression?
- L'accuracy a des limites

. Par exemple, supposons $y_i = [1, 1, 1, 10, 1]$

Cas 1 : $\hat{y}_i = [1, 1, 1, -1, 1]$ donne une accuracy de 80% et un P&L de -6.

Cas 2 : $\hat{y}_i = [-1, 1, 1, 1, 1]$ donne une accuracy de 80% et un P&L de 12.

Cas 3 : $\hat{y}_i = [-1, -1, 1, 1, 1]$ donne une accuracy de 60% et un P&L de 10.





Plan

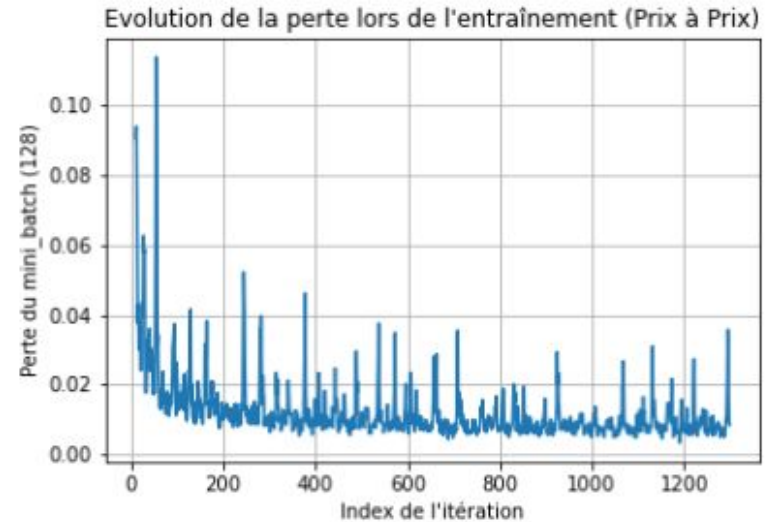
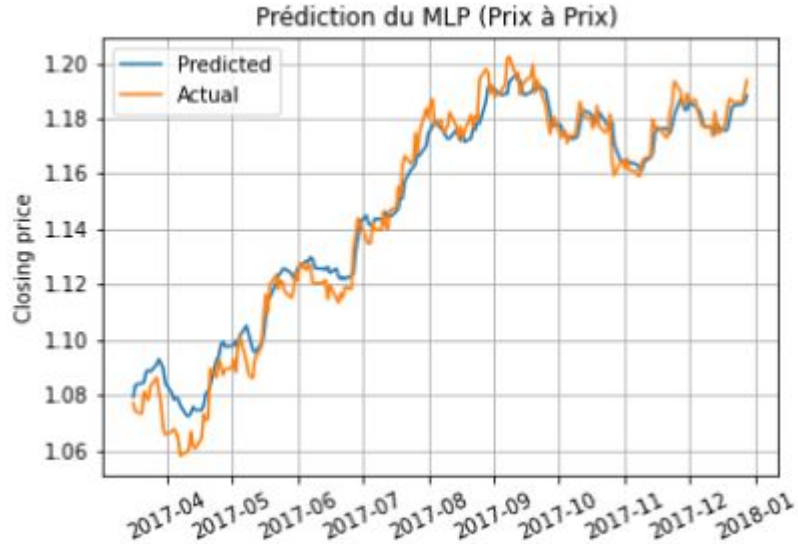
1. Mise en contexte
 - a. Qu'est ce que le trading automatique?
 - b. Passer de la prédiction à la stratégie
 - c. Limite des métriques actuelles
2. Problématique
 - a. Feuille de route
 - b. Ressources à disposition
 - c. Backtesting
3. Coeur du projet
 - a. Modèles de prédiction
 - b. Approche par classification : réfutation des métriques actuelles
 - c. Approche par régression: réfutation des loss function usuelles**
 - d. Métrique et fonction de perte d'exploitabilité réelle
4. Résultats : Comparatif de notre métrique vs. la MSE



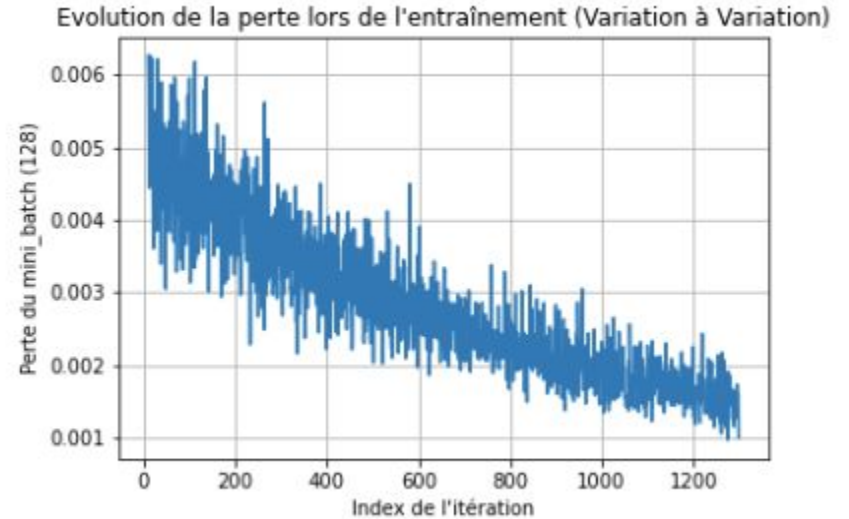
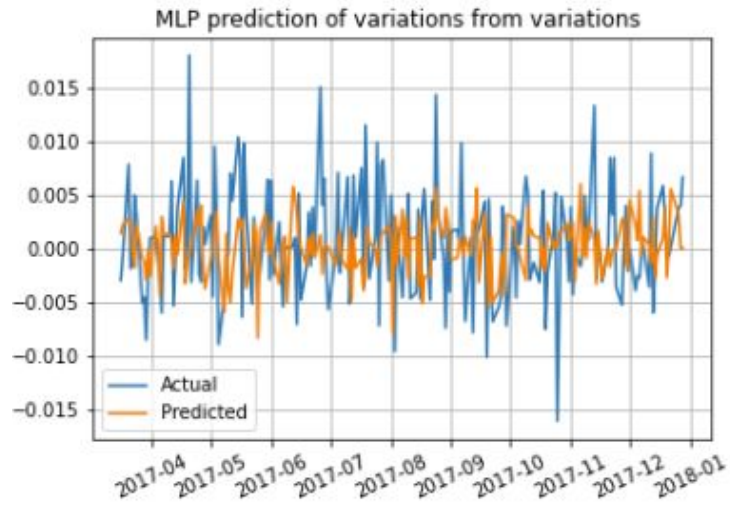
Approche par régression

- Paramètres
 - Output: Amplitude des variations de prix
 - Modèles
 - MLP entraînement sur les prix de fermeture
 - CNN entraînement sur les vecteurs OHLCV
 - Pertes: MSE
- Résultats
 - Prédiction Prix à Prix
 - Prédiction Variation de prix à Variation de prix

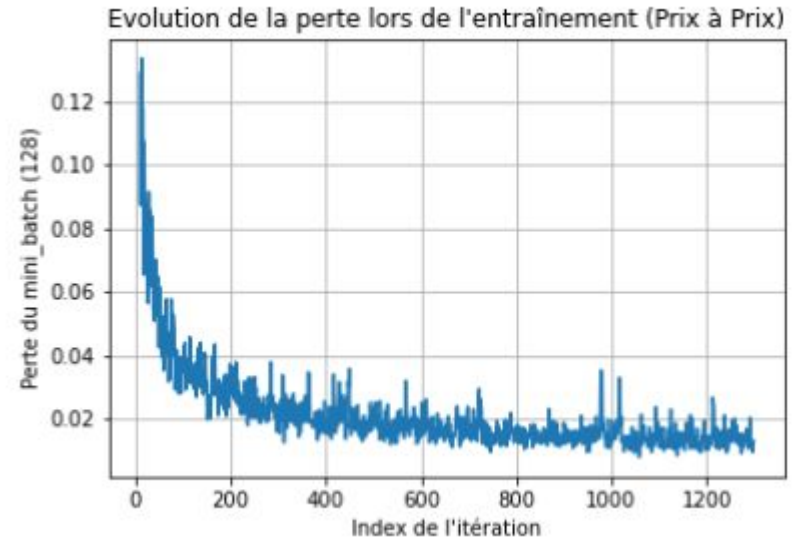
Approche par régression (MLP) Prix



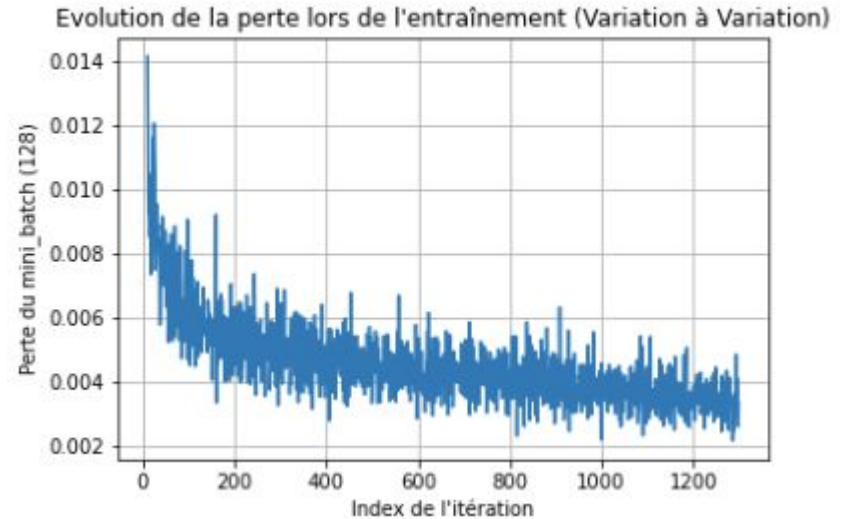
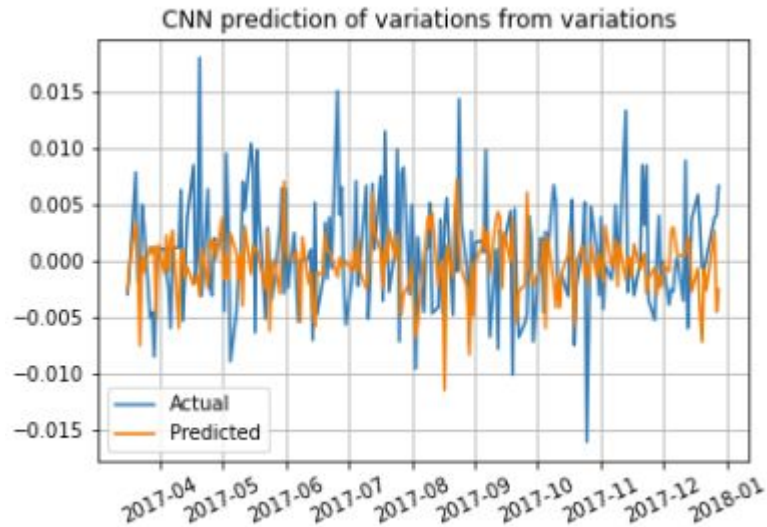
Approche par régression (MLP) Variation



Approche par régression (CNN) Prix



Approche par régression (CNN) Variation





Limites des métriques usuelles en régression

Exemple:

Réel: -1

Prédit: 1 (Prédiction dans le mauvais sens)

MSE: 4

Réel: 10

Prédit: 2 (Prédiction dans le bon sens)

MSE: 64

$$MSE = \frac{1}{n} \sum \underbrace{\left(y - \hat{y} \right)^2}_{\substack{\text{The square of the difference} \\ \text{between actual and} \\ \text{predicted}}}$$



Plan

1. Mise en contexte
 - a. Qu'est ce que le trading automatique?
 - b. Passer de la prédiction à la stratégie
 - c. Limite des métriques actuelles
2. Problématique
 - a. Feuille de route
 - b. Ressources à disposition
 - c. Backtesting
3. Coeur du projet
 - a. Modèles de prédiction
 - b. Approche par classification : réfutation des métriques actuelles
 - c. Approche par régression: réfutation des loss function usuelles
 - d. Métrique et fonction de perte d'exploitabilité réelle**
4. Résultats : Comparatif de notre métrique vs. la MSE



Une métrique corrélée au P&L

- Etape 1: Approche du P&L en classification

$$f_b(\hat{y}, y) = \sum_i q \times \epsilon(y_i \times \hat{y}_i) \times |y_i|$$

TABLEAU 2 – Corrélation de la métrique avec le P&L

Métrique	-125.7	-26.6	-200	-51.5	-75.9
P&L	-125.7	-26.6	-200	-51.5	-75.9
Corrélation	1.0				



Une fonction de perte corrélée au P&L

- Passage à une fonction convexe et minimisable (jusqu'à 0)

$$g(x) = -\ln\left(\frac{1}{(a+1)\alpha}(x + \alpha a)\right)$$

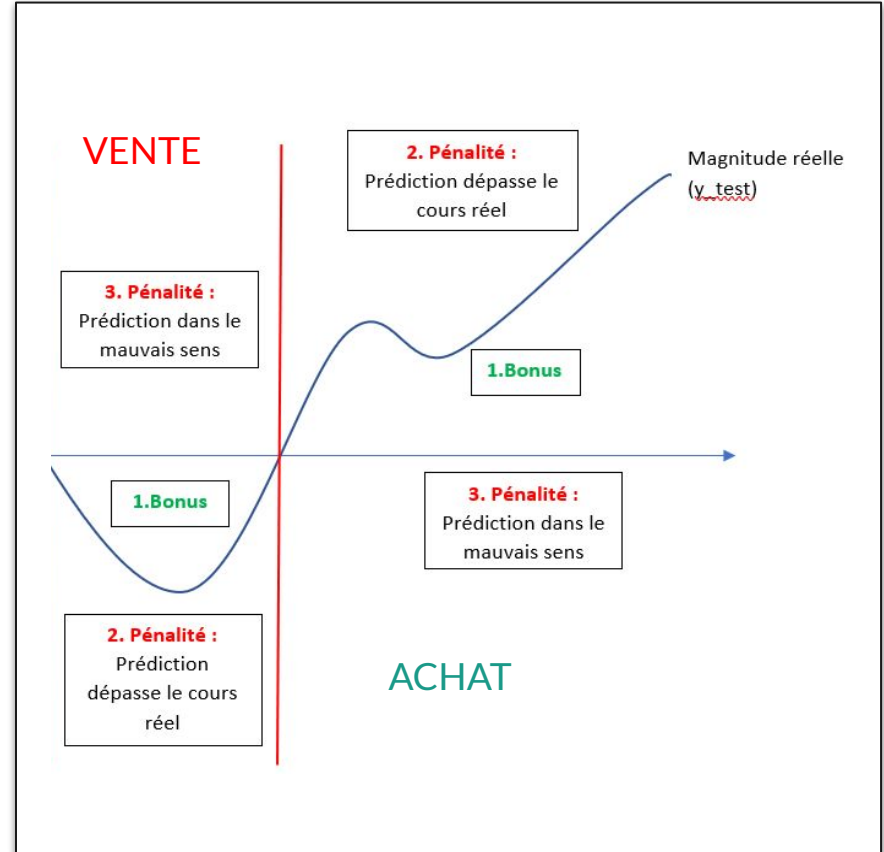
TABLEAU 3 – Corrélation de la fonction de perte avec le P&L

Métrique	0.76	0.71	0.81	0.73	0.72
P&L	-125.7	-26.6	-200	-51.5	-75.9
Corrélation	0.97				

$$\lim_{x \rightarrow -a\alpha} g(x) = +\infty$$

$$g(\alpha) = 0$$

Raffinement de la fonction de perte





Fonction de perte adaptée

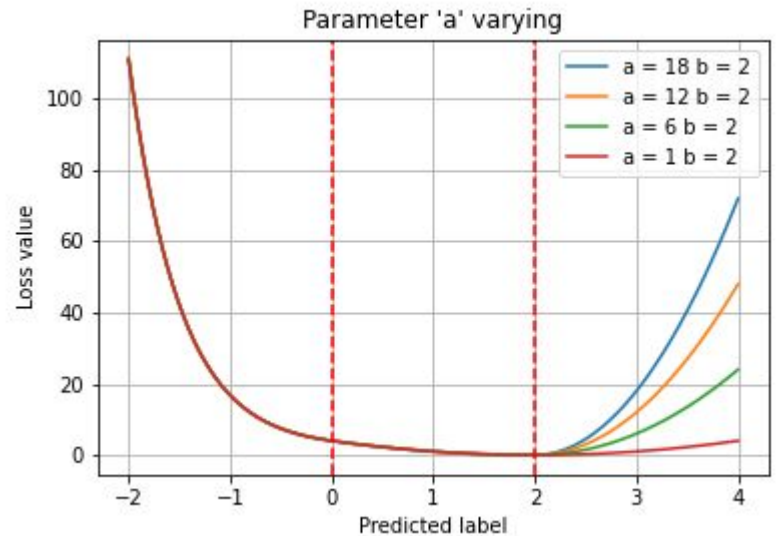
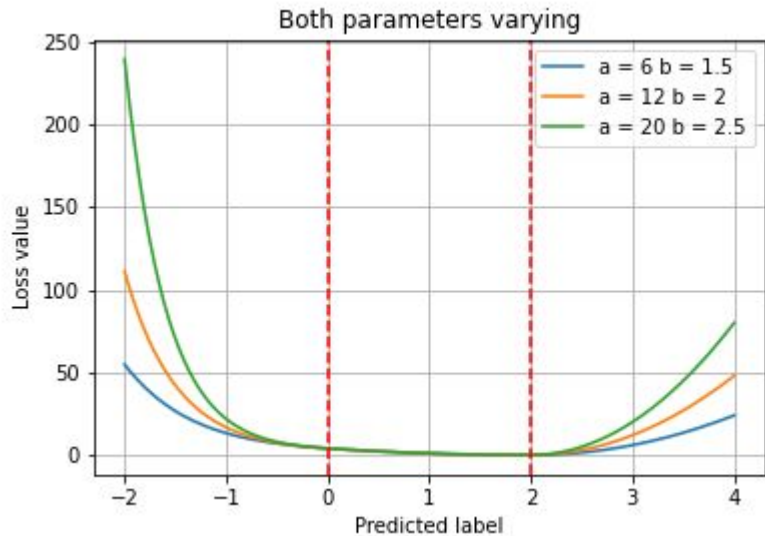
- Passage en régression
- 3 cas de figures sont retranscrits dans une fonction définie par partie avec le comportement voulu
- Loss function pour pouvoir optimiser les algorithmes de deep
- Couple de paramètres (a,b)

$$f_{a,b}(\hat{y}, y) = \frac{2}{b}|\hat{y}|e^{b|y|} + \hat{y}^2 - \frac{2}{b}|\hat{y}| \quad \text{si } c(\hat{y} \times y) < 0$$

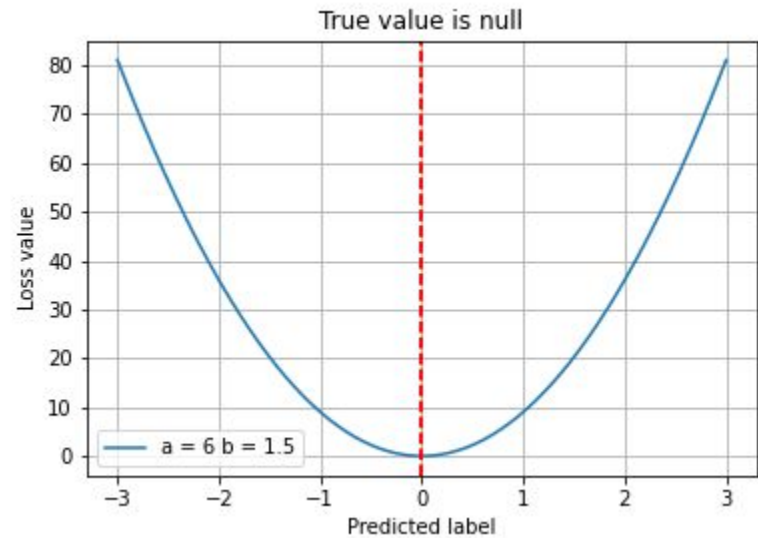
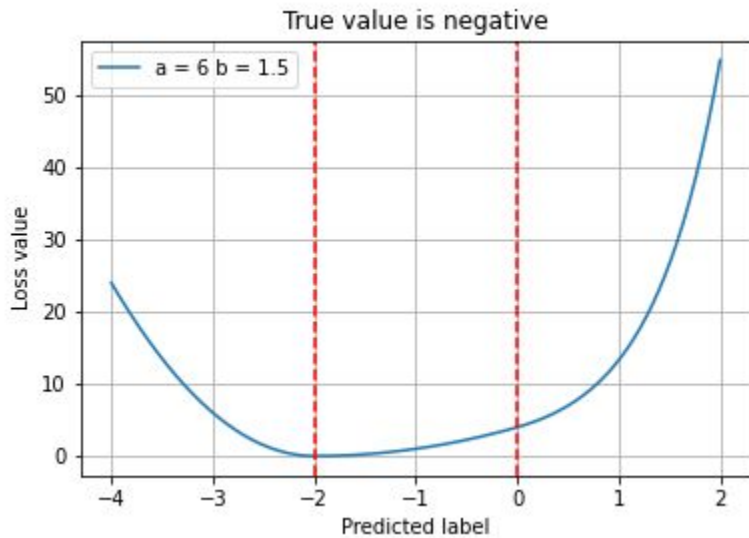
$$f_{a,b}(\hat{y}, y) = (\hat{y} - y)^2 \quad \text{si } c(\hat{y} \times y) \geq 0 \wedge |\hat{y}| < |y|$$

$$f_{a,b}(\hat{y}, y) = a \times (\hat{y} - y)^2 \quad \text{si } c(\hat{y} \times y) \geq 0 \wedge |\hat{y}| \geq |y|$$

Fonction de perte paramétrée



Fonction de perte : cas particuliers





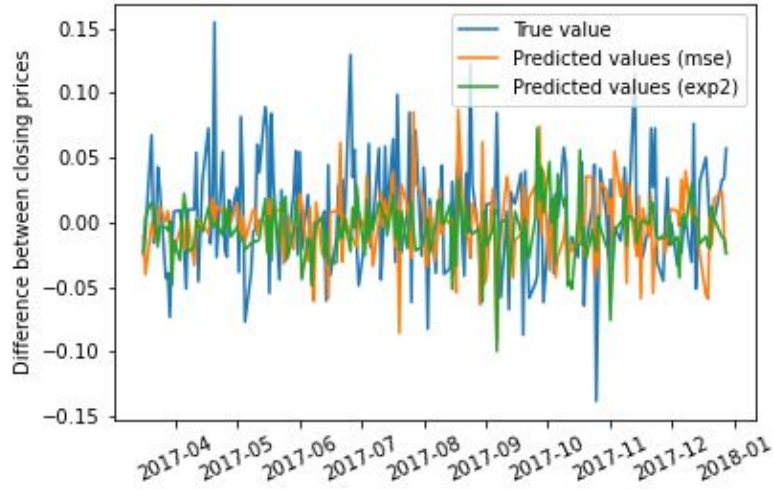
Plan

1. Mise en contexte
 - a. Qu'est ce que le trading automatique?
 - b. Passer de la prédiction à la stratégie
 - c. Limite des métriques actuelles
2. Problématique
 - a. Feuille de route
 - b. Ressources à disposition
 - c. Backtesting
3. Coeur du projet
 - a. Modèles de prédiction
 - b. Approche par classification : réfutation des métriques actuelles
 - c. Approche par régression: réfutation des loss function usuelles
 - d. Métrique et fonction de perte de l'exploitabilité réelle
4. **Résultats : Comparatif de notre métrique vs. la MSE**

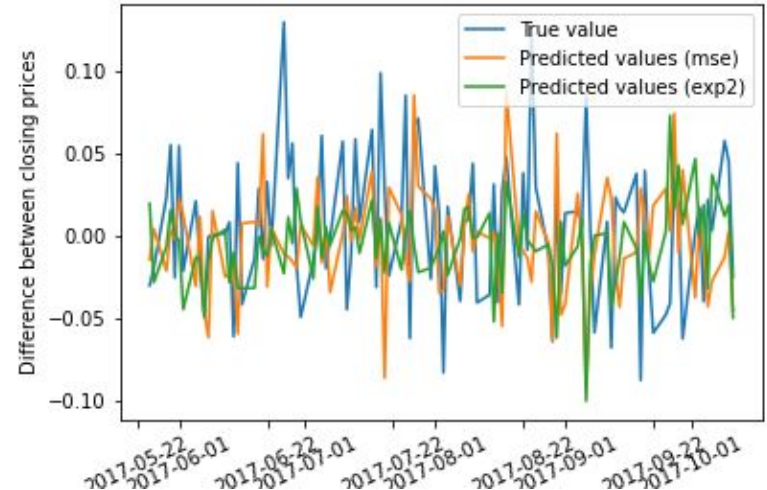
Comparaison avec la MSE

	Corrélation avec le Backtest
MSE	0.11
EXP	0.37

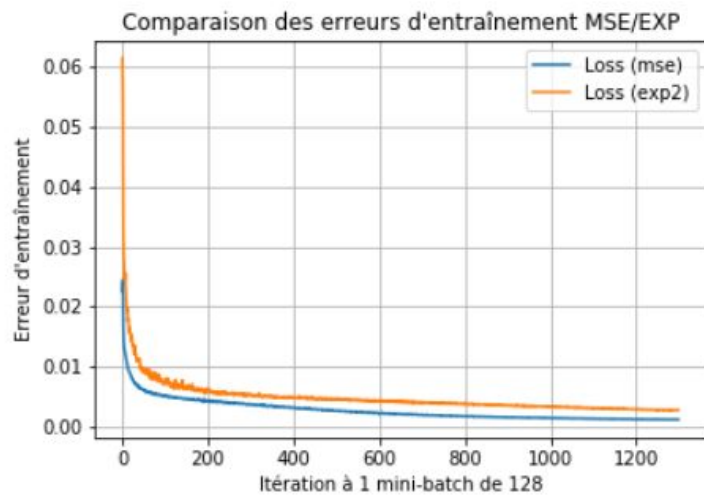
Comparing Predicted and True values on testing dataset (global view)



Comparing Predicted and True values on testing dataset (local view)



Comparaison avec la MSE



Mesure	Métrique	Moyenne	Écart-type	Confiance (95%)
P&L	MSE	-183.8	626.5	88.6
	EXP	-239.5	673.2	95.2
Marge brute	MSE	3998	313.2	44.3
	EXP	3970.2	336.6	47.6
Perte brute	MSE	-4181.9	313.2	44.3
	EXP	-4209.7	336.6	47.6
Facteur de profit	MSE	0.97	0.15	0.02
	EXP	0.96	0.15	0.02
Nombre de trades gagnants	MSE	104.8	5.6	0.8
	EXP	104.06	6.76	0.94
Nombre de trades perdants	MSE	100.24	5.63	0.8
	EXP	100.94	6.66	0.94
Pourcentage rentable	MSE	0.51	0.028	0.004
	EXP	0.5	0.03	0.004
Moy Trade Gain Perte	MSE	-0.89	3.05	0.43
	EXP	-1.17	3.28	0.46
Moy Trade Gagnant	MSE	38.13	2.03	0.29
	EXP	38.15	2.26	0.32
Moy Trade Perdant	MSE	-41.74	2.12	0.3
	EXP	-41.72	2.3	0.32
Ratio Moy Gain Perte	MSE	0.92	0.1	0.01
	EXP	0.92	0.1	0.01
Plus grand trade gagnant	MSE	160	20.5	2.89
	EXP	161	20.8	2.95
Plus grand trade perdant	MSE	-171.5	14.37	2.03
	EXP	-170.3	13.96	1.97



Conclusion

- Projet assez ouvert, beaucoup de rebondissements
- Changement de type de modèles: classifieurs => régresseurs
- Travail important dans la compréhension du problème et la formalisation
- Processus de création de la métrique libre: nous sommes parti du P&L
- Métrique finalement équivalente à la MSE

- Pistes d'amélioration
 - Inclure d'autres types de données pour l'entraînement du modèle (macro-économiques, "market sentiment" ...)
 - Tester de nouvelles stratégies de trading
 - Adopter d'autres angles d'approches pour la création de métrique