

Deep Learning in Finance

TP3 : autoencoders

3 décembre 2024

Le rendu de ce TP est obligatoire

Buts

- Implémenter des autoencoders en mode fonctionnel.
- Analyser les encodages et les séries temporelles générées.

Données

- Télécharger le fichier de données `sp500_etc.csv.gz` afin d'obtenir une matrice (DataFrame) X de dimension $T \times N$. Vous pouvez également choisir une collection de titres de votre choix ($N \geq 100$). Enlever les colonnes des NA.
- conseil : écrire une fonction de construction de modèle qui retourne les différentes parties du modèle, en particulier la partie entrée et le modèle jusqu'au bottleneck, ce qui permet de construire un autre modèle `Model(inputs,bottleneck)` qui peut sortir les états latents.

1 Catégorisation d'actifs

1. Implémenter un autoencoder avec dimension d'étranglement B . Utiliser l'approche fonctionnelle de sorte à pouvoir extraire l'état latent z_i correspondant à une entrée x_i

Séparer les actifs en jeux d'entraînement, de validation et de test.

1.1 Rendements bruts

1. Entraîner un autoencodeur avec les rendements d'actifs $r_i(t)$ pour les i du jeu d'entraînement. Notez que l'indice des lignes dans ce contexte dénote les titres et celui des colonnes celui du temps.
2. Calculer la fraction moyenne de sign $r_i(t)$ apprise correctement par l'AE dans le jeu dans le jeux d'entraînement, de validation et de test.
3. Tracer cette fraction en fonction de B (choisissez les valeurs de B).
4. Extraire les états latents $\{z_i\}$ du jeu de test. Pour $B = 2$, tracer $\{z_{i,2}\}$ en fonction de $\{z_{i,1}\}$. Pour $B = 3$, on peut utiliser un heatmap. Pour $B > 3$, on peut tracer les distributions de chaque composante de z_i ou utiliser un outil de project bidimensionnel comme UMAP.

2 Etats de marché

1. Refaire les points 1.1 à 1.3 mais en entraînant l'AE sur des vecteurs de jours. La catégorisation est donc une catégorisation de jours, ce qui correspond à une détermination d'états de marchés.
2. Comment feriez-vous pour utiliser des AE pour la prédiction ? Comment juger de leur performance ?

3 Détection d'anomalies

1. Tracer l'erreur de reconstruction moyenne des titres en fonction du temps (moyenne journalière). Quels sont les jours qui ont une erreur de reconstruction anormale ? Pouvez-vous lier ces jours à une autre caractéristique du marché ?

2. Calculez l'erreur moyenne de reconstruction des titres (moyenne par titre sur les jours) et classez-les. Quels sont les titres qui ont une erreur de reconstruction anormale ? Que pouvez-vous en dire ?