

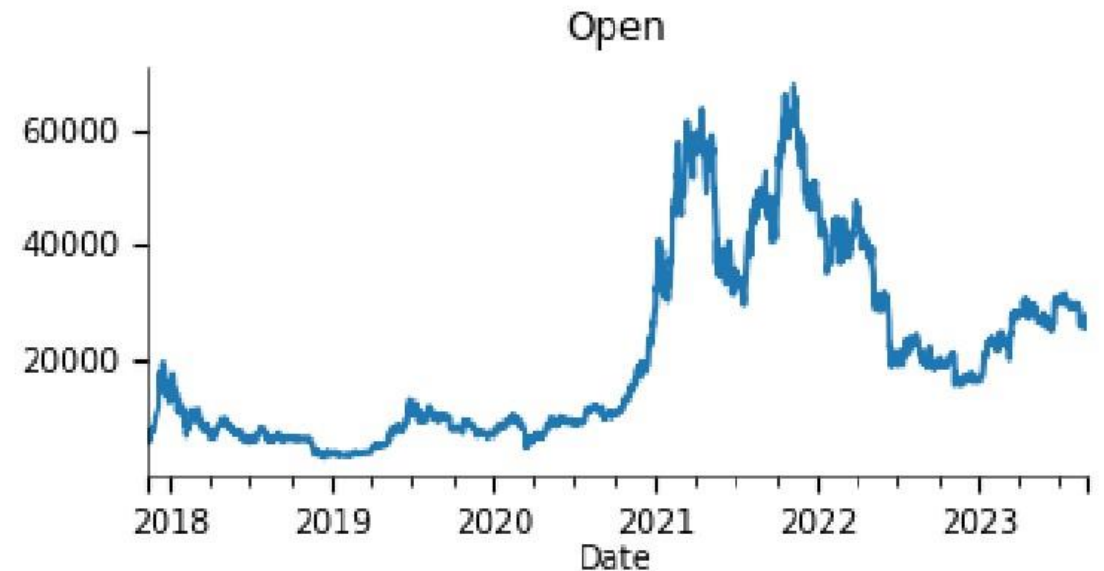
Πρόβλεψη της τιμής των κρυπτονομισμάτων με τεχνικές βαθιάς μάθησης σε δεδομένα μεγάλου όγκου

ΤΡΙΠΟΛΙΤΗΣ ΙΩΑΝΝΗΣ-ΝΙΚΟΛΑΟΣ

AM:1062662

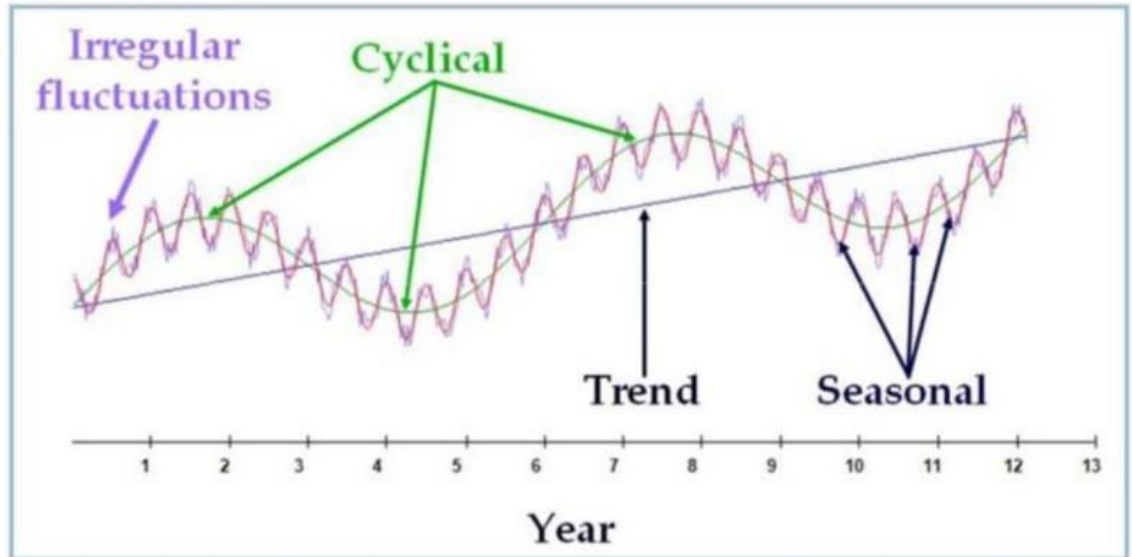
Χαρακτηριστικά χρονοσειρών(1)

► Οι χρονοσειρές ουσιαστικά είναι χρονικά διατεταγμένες παρατηρήσεις X_t καταγεγραμμένες την χρονική στιγμή t . Οι χρονομέτρες μπορούν να διαχωριστούν σε δυο κατηγορίες ανάλογα με την φύση του t , αν οι παρατηρήσεις καταγράφονται σε διακεκριμένες χρονικές στιγμές τότε η χρονοσειρά είναι διακριτή αλλιώς αν καταγράφονται συνεχόμενα σε κάποιο πέρασμα του χρόνου τότε είναι συνεχείς. Στην περίπτωση των κρυπτονομισμάτων προφανώς οι χρονοσειρές θα είναι διακριτές.



Χαρακτηριστικά χρονοσειρών(2)

- ▶ Τάση
- ▶ Εποχικότητα
- ▶ Κυκλικότητα
- ▶ Υπολειπόμενο κομμάτι



Χαρακτηριστικά
χρονοσειρών(3)

Στασιμότητα

- Σταθερό μέσο ορό
- Σταθερή διακύμανση
- Συνδιακύμανση
ανεξάρτητη από τον
χρόνο

Μεγάλα δεδομένα σε χρονοσειρές

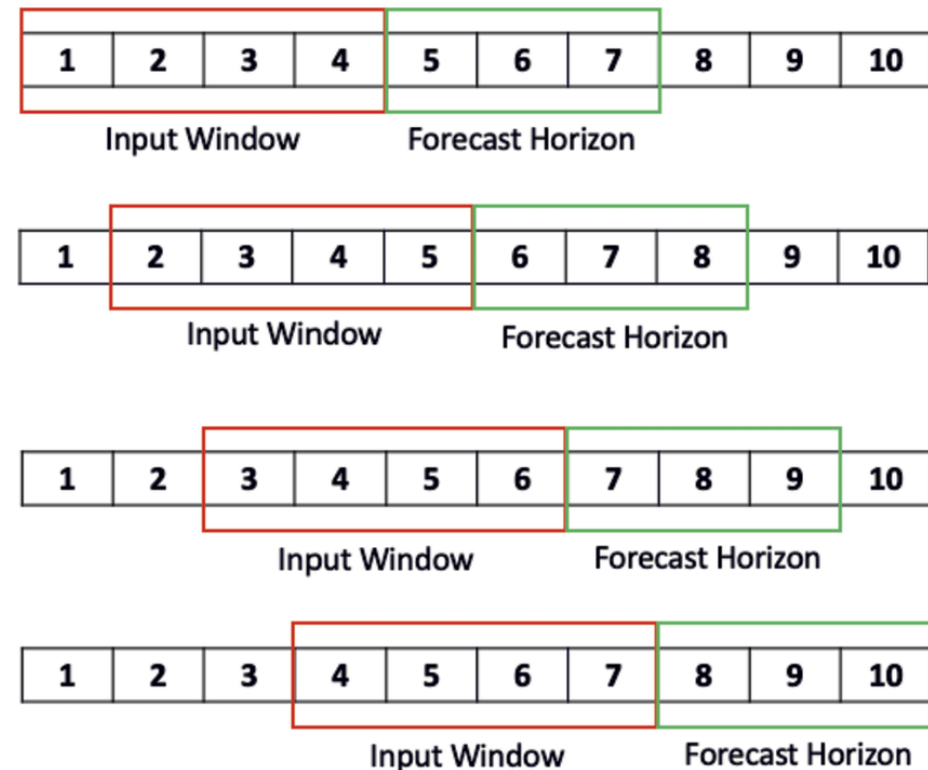
Λαμβάνεις υπόψιν δεδομένα πολύ
πίσω στο χρόνο

Πολύ μικρό χρονικό βήμα

Μεγάλο πλήθος ανεξάρτητων
μεταβλητών για την πρόβλεψη της
εξαρτώμενης μεταβλητής

Η τεχνική της ολίσθησης παραθύρου

► Τα δεδομένα χρονοσειρών μπορούν να παραφραστούν ως δεδομένα εποπτευομένης μάθησης. Αυτό μπορεί να επιτευχθεί με τον μετασχηματισμό των δεδομένων ώστε να παρομοιαστούν ως πρόβλημα εποπτευομένης μάθησης αν μας έχει δοθεί μια ακολουθία αριθμών μιας χρονοσειράς. Αυτό μπορεί να επιτευχθεί χρησιμοποιώντας τα προηγούμενα χρονικά βήματα ως μεταβλητές εισόδου και χρησιμοποιώντας το επόμενο χρονικό βήμα ως μεταβλητή εξόδου.



Στατιστική ανάλυση

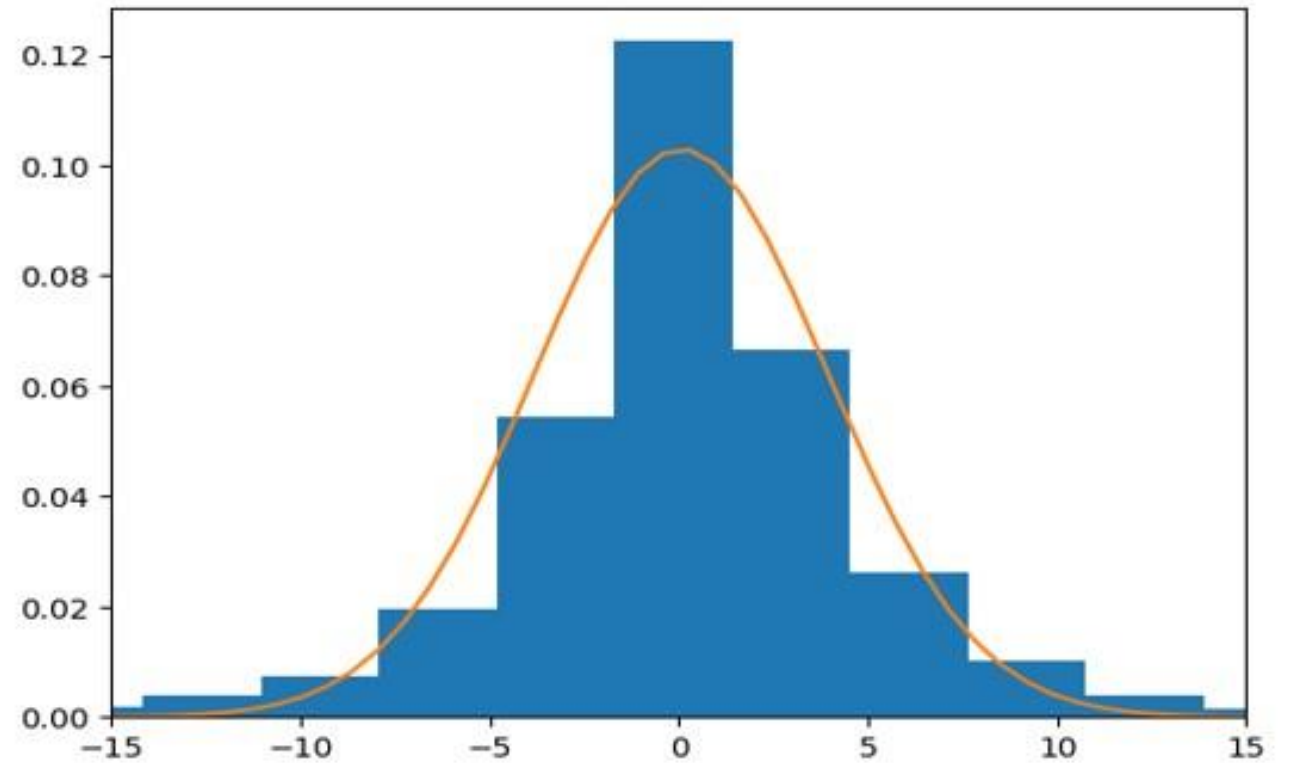
ACF → Αυτοσυσχέτιση

PACF → Μερικώς
αυτοσυσχέτιση

Dickey Fuller →
Στασιμότητα

Αστάθεια

► Για την παρατήρηση της αστάθειας της χρονοσειράς που μελετάμε συστήνεται η οπτικοποίηση του rate of return



Επισκόπηση πλαισίου

Είσοδοι vs. Έξοδοι

Ενδογενής vs. Εξογενής

Αδόμητη vs Δομημένη

Παλινδρόμηση vs Ταξινόμηση

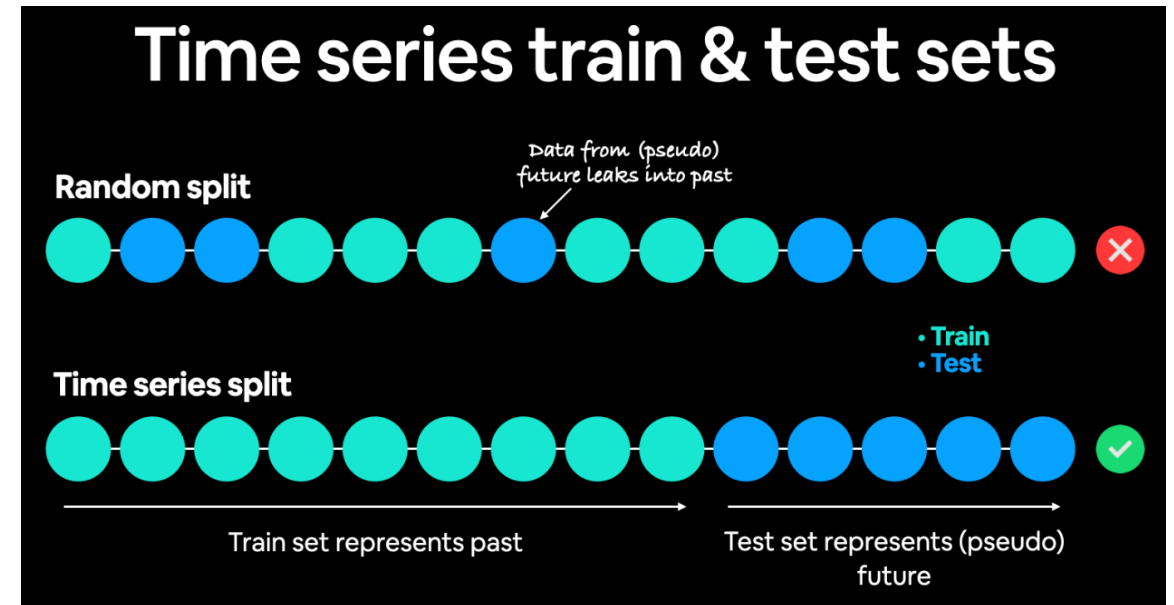
Μονομεταβλητή vs Πολυμεταβλητή

Μονοβηματική vs Πολυβηματική

Στατική vs Δυναμική

Διαχωρισμός δεδομένων

► Είναι απαραίτητο να διατηρηθεί η σειρά των δεδομένων σε μια χρονοσειρά συνεπώς πρέπει να γίνει ένας αυθαίρετος διαχωρισμός των δεδομένων πριν την εκπαίδευση για το παρελθόν, παρόν και μέλλον. Δεν μπορεί να ακολουθηθεί η τεχνική του ανακατέματος των δεδομένων πριν την εκπαίδευση όπως συνηθίζεται διότι τότε παρέχουμε στο μοντέλο μας πληροφορίες για το μέλλον τις οποίες δεν θα έπρεπε να έχει στην διάθεση του.



Στρατηγικές πρόβλεψεις(1)

Αφελής πρόβλεψη

Οικογένεια Arima

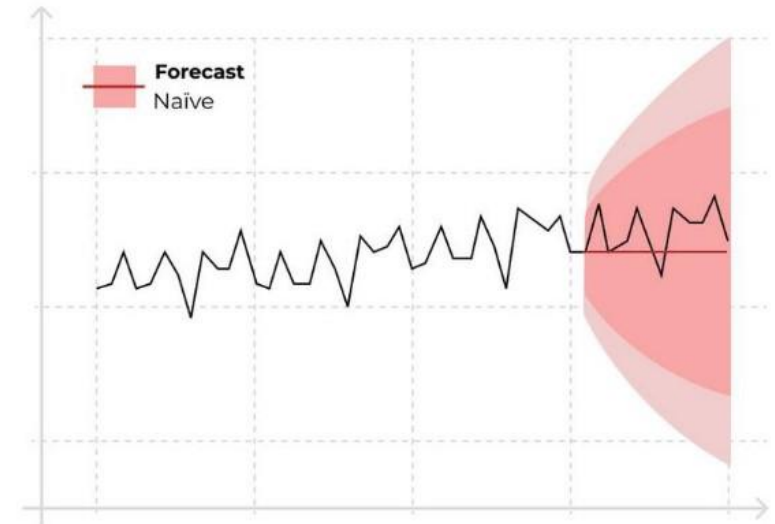
Εκθετική εξομάλυνση

Νευρωνικά δίκτυα

Στρατηγικές πρόβλεψεις(2)

Αφελής πρόβλεψη : Το αφελές μοντέλο χρησιμοποιείται συνήθως ως η γραμμή αναφοράς και δεν χρειάζεται καμία απολύτως εκπαίδευση, αυτό συμβαίνει διότι στην πραγματικότητα το μόνο πράγμα το οποίο κάνει είναι να θεωρεί πως η τιμή στο επόμενο χρονικό βήμα θα είναι ακριβώς η ίδια με την τιμή του προηγούμενου χρονικού βήματος. Μπορεί να ακούγεται απλό αλλά σε ένα ανοιχτό σύστημα όπως για παράδειγμα το χρηματιστήριο είναι ιδιαίτερα δύσκολο να πέτυχει καλύτερη πρόβλεψη από αυτή του αφελούς μοντέλου.

NAÏVE MODEL



Στρατηγικές
πρόβλεψεις(3)

Οικογένεια Arima

- AR → Autoregressive
- I → Integrated
- MA → Moving average
- S → Seasonal
- X → Exogenous

Στρατηγικές πρόβλεψεις(4)

- ▶ Εκθετική εξομάλυνση: Η κυρίως ιδέα είναι να δοθεί μεγαλύτερη σημασία στις πιο πρόσφατες παρατηρήσεις, συνεπώς όσο παλιώνουν οι παρατηρήσεις η σημασία τους γίνεται εκθετικά χαμηλότερη
- ▶ Υπάρχουν τρεις κυρίως μέθοδοι εκθετική εξομάλυνσης
 - ▶ Μονή: 1 παράμετρο E όσο μεγαλύτερο τόσο πιο πολύ εστιάζει σε πρόσφατες τιμές
 - ▶ Διπλή: Προστίθεται 2 παράμετρο για την τάση
 - ▶ Τριπλή: Προστίθεται 3 παράμετρο για την εποχικότητα
- ▶ Όλες οι παράμετροι έχουν τιμές $[0,1]$

Στρατηγικές πρόβλεψεις(5)

► **Νευρωνικά δίκτυα:** Μπορούν να σχεδιαστούν με διάφορους τρόπους όπως MLP,RNN,LSTM. Τα κυρία πλεονεκτήματα των βαθιών νευρωνικών δικτύων σε σχέση με τοις υπόλοιπες μεθόδους προβλέψεις είναι πως δεν χρειάζεται να γίνει κάποια υπόθεση για την εφαρμογή τους και έχουν την ικανότητα να εντοπίσουν σημαντικές γενικευτικές δυνάμεις

MLP

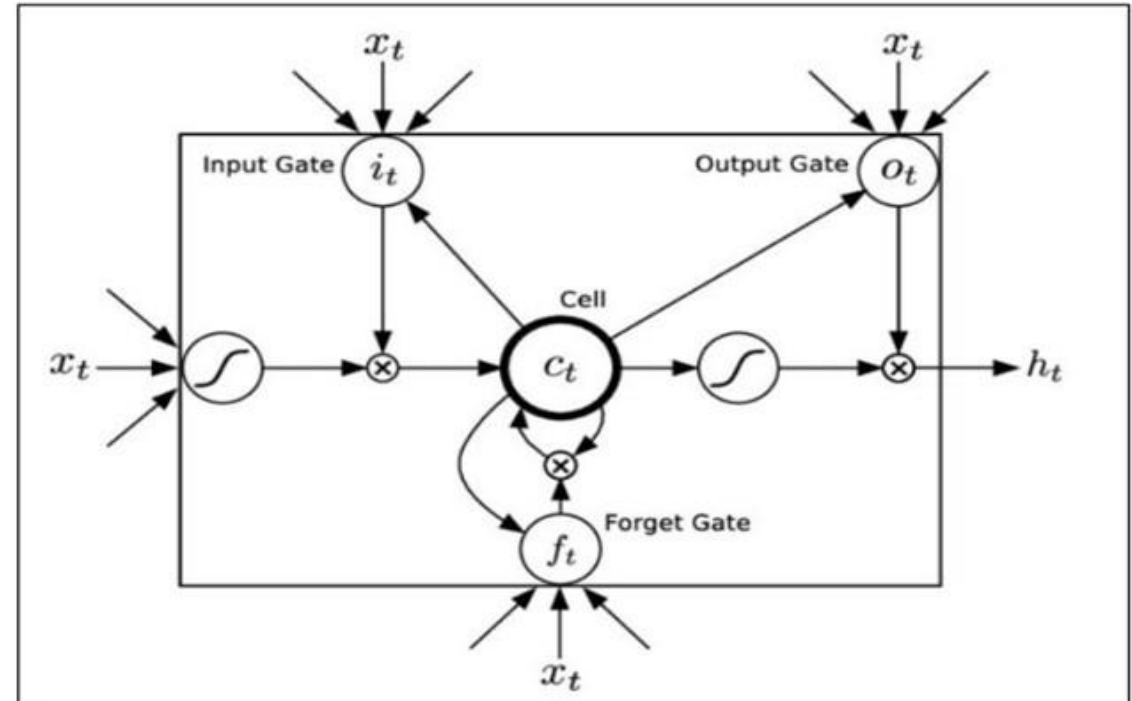
- ▶ Ουσιαστικά δημιουργούν μια συνάρτηση mapping από τις εισόδους στις εξόδους. Αυτή τους η ιδιότητα τα κάνει πολύτιμα για τους εξής λόγους.
- ▶ Ανθεκτικά στην φασαρία και ικανά να διαχειριστούν ακόμα και κενές τιμές
- ▶ Μη γραμμικά
- ▶ Μπορούν να διαχειριστούν πολλές εισόδους και να παράγουν πολυβηματικές εξόδους
- ▶ Όμως έχουν και κάποια μειονεκτήματα λόγω της λειτουργίας τους
- ▶ Πρέπει να είναι προκαθορισμένα πριν την εκπαίδευση τους το μέγεθος της εισόδου(lag) και τα βήματα της εξόδου τα οποία χωρίς στατιστική ανάλυση δεν είναι δυνατό να γνωρίζονται

RNN

► Έχουν έμφυτη υποστήριξη για δεδομένα εισόδου διαδοχικών παρατηρήσεων και παρουσιάζουν όλα τα πλεονεκτήματα των MLP και λύνουν τα προβλήματα της προκαθορισμένης εισόδου καθώς μπορούν να αλλάξουν δυναμικά από μια αρκετά μεγάλη είσοδο διαδοχικών δεδομένων παρουσιάζουν όμως και αυτά το πρόβλημα των μακροχρόνιων εξαρτήσεων. Αυτό εμφανίζεται στα RNN διότι έχουν επαναλαμβανόμενες συνδέσεις οι οποίες τους επιτρέπουν να αποθηκεύουν πληροφορίες από προηγούμενα χρονικά βήματα και είναι υπεύθυνο για τα προβλήματα της εξαφανιζόμενης η εκρηκτικής κλίσης τα οποία εμφανίζονται όταν η κλίση γίνεται πολύ μεγάλη ή πολύ μικρή κατά την εκτέλεση του αλγορίθμου πίσω διαδρομής

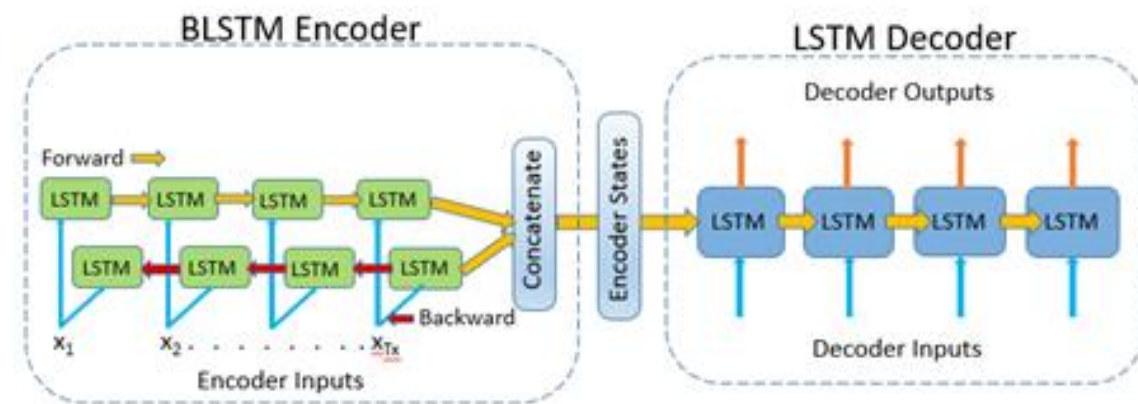
LSTM

- ▶ Είναι μια κατηγορία των RNN και ειδικεύεται στην εξαφάνιση του προβλήματος των μακροχρονίων εξαρτήσεων. Έχει ταυτόχρονα μακροπρόθεσμη και βραχυπρόθεσμη μνήμη και αποτελείται από τρία κομμάτια κάθε ένα υπεύθυνο για διαφορετική λειτουργία.
- ▶ υπεύθυνο για να επιλέξει αν η πληροφορία που έρχεται αξίζει να απομνημονευθεί ή είναι αδιάφορη
- ▶ υπεύθυνο για να μάθει την πληροφορία εισόδου
- ▶ υπεύθυνο για να εκσυγχρονίσει την πληροφορία από το ένα χρονικό βήμα στο επόμενο



Bidirectional LSTM

► Τα LSTM διπλής κατεύθυνσης είναι και αυτά ένα είδος RNN, τα οποία έχουν την ιδιαιτερότητα ότι επεξεργάζονται δεδομένα από δυο κατεύθυνσης, από εμπρός και από πίσω. Στα παραδοσιακά LSTM η πληροφορία κυλάει από το παρελθόν προς το μέλλον και κάνει προβλέψεις με αυτόν τον τρόπο. Όμως στα LSTM διπλής κατεύθυνσης το δίκτυο λαμβάνει υπόψιν και το μέλλον το οποίο του επιτρέπει να καταλαβαίνει εξαρτήσεις και από τις δυο κατεύθυνσης. Αποτελείται από δυο LSTM το ένα λαμβάνει την είσοδο από εμπρός και το άλλο από πίσω, αυτό του δίνει πληροφορίες για τα παρόντα αλλά και για τα μελλοντικά χρονικά βήματα.



Εφαρμογή(1)

Έμπνευση...

- ▶ Μάθημα <Συστήματα Διαχείρισης Μεγάλων Δεδομένων> (CEID1176) του τμήματος Μηχανικών Η/Υ και Πληροφορικής
- ▶ Εργασία 2021-2022 και το κεφάλαιο για time series

Εφαρμογή(2)

Τεχνολογίες...

- ▶ Τα δεδομένα συγκεντρώθηκαν από το Yahoo Finance
- ▶ Τα βαθεία νευρωνικά δίκτυα υλοποιήθηκαν με tensorflow
- ▶ Ο κώδικας εκτελείται στο cloud στο google colaboratory