

Λύση θέματος χρονοσειρών

Η χρονοσειρά μου είναι η εξής

| | | | | | | | | | | | |
|------|---|---|---|---|---|---|---|---|---|----|----|
| t | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| y(t) | 1 | 2 | 2 | 3 | 3 | 4 | 4 | 4 | 5 | 5 | 5 |

Οι **προβλέψεις** των σημείων 8...11 με βάση τα τρία μοντέλα είναι οι εξής

| | | | | | | | | | | | |
|------------|--|--|--|--|--|--|--|-------------|-------------|-------------|-------------|
| naive | | | | | | | | 4 | 4 | 4 | 4 |
| mean | | | | | | | | 2.71 | 2.71 | 2.71 | 2.71 |
| regression | | | | | | | | 3.52 | 3.84 | 4.16 | 4.49 |

Οι **δείκτες** MAE/ MAPE της ακρίβειας των τριών μοντέλων είναι

| | | | | | | | | | | | | |
|-----------------|--|--|--|--|--|--|--|--------|--------|--------|--------|---------------|
| MAE NAIVE | | | | | | | | 0 | 1 | 1 | 1 | 0.750 |
| MAE MEAN | | | | | | | | 1.29 | 2.29 | 2.29 | 2.29 | 2.036 |
| MAE REGRESSION | | | | | | | | 0.48 | 1.16 | 0.84 | 0.51 | 0.748 |
| MAPE NAIVE | | | | | | | | 0.00% | 20.00% | 20.00% | 20.00% | 15.00% |
| MAPE MEAN | | | | | | | | 32.14% | 45.71% | 45.71% | 45.71% | 42.32% |
| MAPE REGRESSION | | | | | | | | 12.05% | 23.18% | 16.72% | 10.26% | 15.55% |

A. Η γραμμική παλινδρόμηση έχει τον καλύτερο (μικρότερο) MAE δείκτη ενώ το απλοϊκό μοντέλο έχει τον καλύτερο δείκτη MAPE. Το μοντέλο του μέσου όρου είναι πολύ χειρότερο προφανώς διότι αποτυγχάνει να ‘δει’ την αυξανόμενη τάση των δεδομένων. Το ίδιο κακό θα ήταν και το απλοϊκό αλλά στάθηκε ‘τυχερό’ γιατί τη στιγμή 8 η χρονοσειρά είχε την ίδια τιμή με τη στιγμή 7, κάτι που το ‘βοήθησε’ να ξεκινήσει από μηδενικό σφάλμα.

Γ. Αν χρησιμοποιήσω όλα τα δεδομένα για να κατασκευάσω μοντέλο γραμμικής παλινδρόμησης, χωρίς δηλαδή διάκριση train/ test τότε η πρόβλεψη και οι δείκτες για το μοντέλο παλινδρόμησης θα είναι

| | | | | | | | | | | | | |
|--------------------|--|--|--|--|--|--|--|-------------|-------------|-------------|-------------|---------------|
| REGRESSION 11 | | | | | | | | 3.51 | 3.86 | 4.21 | 4.56 | |
| MAE REGRESSION 11 | | | | | | | | 0.49 | 1.14 | 0.79 | 0.45 | 0.717 |
| MAPE REGRESSION 11 | | | | | | | | 12.23% | 22.82% | 15.86% | 8.90% | 14.95% |

Δ. Παρατηρώ ότι η ακρίβεια είναι ελαφρά βελτιωμένη, πράγμα που οφείλεται στο γεγονός ότι το συγκεκριμένο μοντέλο κατασκευάστηκε χρησιμοποιώντας τα δεδομένα που κλήθηκε να προβλέψει, κάτι που ασφαλώς οδηγεί σε καλύτερες ακρίβειες αλλά είναι μεθοδολογικά απόλυτα εσφαλμένο.