

# Περιεχόμενα

Πρόλογος .....	11
<b>Κεφάλαιο 1:</b> Νευρωνικά Δίκτυα και Χρηματοοικονομική Καινοτομία .....	15
1.1 Εισαγωγή.....	15
1.2 Χρηματοοικονομική Μηχανική .....	17
1.3 Συγγενή Γνωστικά Πεδία.....	19
1.4 Οι Λειτουργίες της Χρηματοοικονομικής Μηχανικής.....	21
1.5 Τύποι Εφαρμογών Νευρωνικών Συστημάτων .....	24
1.6 Εφαρμογές Νευρωνικών Συστημάτων στη Χρηματοοικονομική Μηχανική .....	26
1.7 Συμπεράσματα .....	27
Βιβλιογραφία.....	28
<b>Κεφάλαιο 2:</b> Θεμελιώδεις Έννοιες και Βασικά Μοντέλα Νευρωνικών Δικτύων	31
2.1 Εισαγωγή.....	31
2.2 Σύντομη Ιστορική Αναδρομή.....	32
2.3 Ο Βιολογικός Νευρώνας .....	34
2.4 Ο Τεχνητός Νευρώνας McCullock και Pitts .....	36
2.5 Το Γενικό Πλαίσιο της Παράλληλης Κατανεμημένης Επεξεργασίας ...	40
2.6 Δίκτυα με Συναρτήσεις Ενεργοποίησης Τύπου Κατωφλίου.....	49
2.7 Perceptron .....	51
2.8 Δίκτυα Adaline .....	54
2.9 Ο Κανόνας Δέλτα.....	55
2.10 Συμπεράσματα .....	59
Βιβλιογραφία.....	59
<b>Κεφάλαιο 3:</b> Νευρωνικά Δίκτυα Οπισθοδιάδοσης .....	63
3.1 Εισαγωγή.....	63
3.2 Δίκτυα Προτροφοδότησης με Πολλαπλά Στρώματα .....	63
3.3 Ο Γενικευμένος Κανόνας Δέλτα .....	65

3.4	Η Οπισθοδιάδοση στην Πράξη.....	71
3.5	Γεωμετρική Ερμηνεία της Οπισθοδιάδοσης.....	75
3.6	Παράλυση Δικτύου.....	79
3.7	Τοπικά Ελάχιστα.....	82
3.8	Μη Μοναδικές Λύσεις.....	85
3.9	Επεκτάσεις του Αλγορίθμου Οπισθοδιάδοσης.....	86
3.10	Συμπεράσματα.....	87
	Βιβλιογραφία.....	87
<b>Κεφάλαιο 4: Μη Παραμετρική Εκτίμηση με Νευρωνικά Δίκτυα.....</b>		<b>89</b>
4.1	Εισαγωγή.....	89
4.2	Διαμόρφωση Αναφοράς.....	89
4.3	Η Στατιστική Θεώρηση της Νευρωνικής Μάθησης.....	92
4.4	Η Διαδικασία Εκτίμησης Νευρωνικών Υποδειγμάτων.....	102
4.5	Συμπεράσματα.....	107
	Βιβλιογραφία.....	107
<b>Κεφάλαιο 5: Η Διαδικασία Εκτίμησης Ενός Νευρωνικού Μοντέλου.....</b>		<b>109</b>
5.1	Εισαγωγή.....	109
5.2	Ο Κίνδυνος Πρόβλεψης.....	114
5.3	Αλγεβρική Εκτίμηση του Κινδύνου Πρόβλεψης.....	116
5.4	Η Εκτίμηση του Μητρώου Συνδιακύμανσης.....	119
5.5	Εκτίμηση του Κινδύνου Πρόβλεψης με Μεθόδους Επαναλαμβανόμενης Δειγματοληψίας.....	121
5.6	Διαγνωστικός Έλεγχος.....	124
5.7	Έλεγχος για σειριακή συσχέτιση στα κατάλοιπα.....	125
5.8	Ένας F-έλεγχος για την επάρκεια του υποδείγματος.....	127
5.9	Η Επιλογή Νευρωνικού Μοντέλου στην Πράξη.....	128
5.10	Συμπεράσματα.....	133
	Βιβλιογραφία.....	134
<b>Κεφάλαιο 6: Διαστήματα Εμπιστοσύνης και Πρόβλεψης.....</b>		<b>137</b>
6.1	Εισαγωγή.....	137
6.2	Διαστήματα Εμπιστοσύνης Έναντι Διαστημάτων Πρόβλεψης.....	138
6.3	Προσεγγίσεις του Νευρωνικού μοντέλου με Επεκτάσεις Taylor.....	142

6.4	Μέθοδοι Μέγιστης Πιθανοφάνειας για την Εκτίμηση Διακύμανσης Θορύβου Εξαρτώμενης από το Διάνυσμα Εισόδου.....	148
6.5	Σύνολα Νευρωνικών Δικτύων.....	150
6.6	Συμπεράσματα.....	154
	Βιβλιογραφία.....	155
<b>Κεφάλαιο 7:</b>	<b>Εκτίμηση Αποδόσεων, Μεταβλητοτήτων και Συσχετίσεων.....</b>	<b>157</b>
7.1	Εισαγωγή.....	157
7.2	Η Διαδικασία Κατασκευής Χαρτοφυλακίου.....	158
7.3	Εκτιμήσεις Αποδόσεων.....	167
7.4	Η Ευαισθησία των Αποδόσεων στους Παράγοντες Κινδύνου.....	171
7.5	Εκτιμήσεις Διακυμάνσεων και Συσχετίσεων.....	173
7.6	Συμπεράσματα.....	180
	Βιβλιογραφία.....	180
<b>Κεφάλαιο 8:</b>	<b>Προεπεξεργασία Δεδομένων και Αρχική Επιλογή των Παραμέτρων του Δικτύου και του Αλγορίθμου Εκπαίδευσης.....</b>	<b>183</b>
8.1	Εισαγωγή.....	183
8.2	Αρχική Επιλογή Μεταβλητών και Συλλογή Δεδομένων.....	184
8.3	«Καθάρισμα» Δεδομένων.....	186
8.4	Παρατηρήσεις που Ασκούν Υπέρμετρη Επιρροή στο Μοντέλο.....	188
8.5	Μη Στάσιμες Μεταβλητές.....	190
8.6	Εξομάλυνση Δεδομένων ή Φιλτράρισμα Θορύβου.....	197
8.7	Φιλτράρισμα Δεδομένων.....	198
8.8	Αντικρουόμενα Δεδομένα.....	199
8.9	Αλλαγή Κλίμακας.....	199
8.10	Δημιουργία Συνόλων Εκπαίδευσης, Ελέγχου και Επαλήθευσης.....	202
8.11	Επιλογή Παραμέτρων του Δικτύου και του Αλγορίθμου Εκπαίδευσης.....	205
8.12	Συμπεράσματα.....	208
	Βιβλιογραφία.....	208
<b>Κεφάλαιο 9:</b>	<b>Εφαρμογές Πρόβλεψης Χρονολογικών Σειρών και Ταξινόμησης.....</b>	<b>211</b>
9.1	Εισαγωγή.....	211
9.2	Αποδόσεις Χρονολογικών Σειρών.....	214

9.3	Μη Δεσμευμένες Κατανομές Αποδόσεων.....	220
9.4	Πρόβλεψη Τιμών ή Αποδόσεων; .....	223
9.5	Έλεγχοι Κανονικότητας.....	224
9.6	Είναι οι Αποδόσεις Στάσιμες;.....	228
9.7	Είναι οι Αποδόσεις Προβλέψιμες;.....	231
9.8	Εφαρμογές Ταξινόμησης .....	238
9.9	Υποθέσεις και Αντικειμενικοί Στόχοι της Διακριτής Ανάλυσης .....	241
9.10	Ο Πίνακας Ταξινόμησης.....	242
9.11	Νευρωνικά Μοντέλα Ταξινόμησης .....	246
9.12	Συμπεράσματα .....	248
	Βιβλιογραφία.....	249
<b>Κεφάλαιο 10: Αξιολόγηση Μοντέλου .....</b>		<b>251</b>
10.1	Εισαγωγή.....	251
10.2	Κριτήρια Αξιολόγησης της Ικανότητας Πρόβλεψης του Δικτύου .....	254
10.3	Αποδόσεις και Στρατηγικές .....	268
10.4	Κριτήρια Αξιολόγησης Στρατηγικών Συναλλαγών.....	279
10.5	Κριτήρια Αξιολόγησης της Ικανότητας Ταξινόμησης του Δικτύου .....	282
10.6	Συμπεράσματα .....	286
	Βιβλιογραφία.....	286
<b>Κεφάλαιο 11: Υφιστάμενες Τεχνολογίες και Συστήματα .....</b>		<b>289</b>
11.1	Εισαγωγή.....	289
11.2	Ward Systems .....	290
11.3	BrainMaker Standard & Professional.....	294
11.4	NeuroDimension, TradingSolutions .....	296
11.5	NeuralWare, NeuralWorks Professional II/PLUS v5.5 .....	300
11.6	Vesta Services, Qnet v2000 .....	301
11.7	NeuroXL .....	302
11.8	Attrasoft PredictorPro .....	302
11.9	MathWorks MATLAB, Neural Network Toolbox 4.0.2.....	303
11.10	Συμπεράσματα .....	303

## Κεφάλαιο 5

# Η Διαδικασία Εκτίμησης Ενός Νευρωνικού Μοντέλου

### 5.1 Εισαγωγή

Σ' αυτό το κεφάλαιο περιγράφεται η διαδικασία εκτίμησης ενός νευρωνικού μοντέλου, η οποία είναι αρκετά πιο περίπλοκη από την αντίστοιχη διαδικασία για τους παραμετρικούς εκτιμητές. Τα νευρωνικά δίκτυα είναι μία πολύ σημαντική κατηγορία μοντέλων εξαιτίας της ιδιότητας «καθολικής προσέγγισης» που τα χαρακτηρίζει. Το κόστος αυτής της ευελιξίας όμως, είναι το πρόβλημα του ελέγχου της πολυπλοκότητας της αρχιτεκτονικής του δικτύου, που υπό το πρίσμα της στατιστικής θεωρίας αντιστοιχεί στο πρόβλημα της εξισορρόπησης της μεροληψίας και της διακύμανσης της συνάρτησης τετραγωνικού σφάλματος.

Η παρουσία πολλών ελεύθερων παραμέτρων κάνουν τα νευρωνικά δίκτυα δύσκολα στην ανάλυση και την ερμηνεία. Από την άλλη, τα μοντέλα γραμμικής παλινδρόμησης είναι πολύ απλούστερα στην κατασκευή και την κατανόηση, αλλά σε πολλές περιπτώσεις αποδεικνύονται ανεπαρκή για την προσέγγιση περίπλοκων διαδικασιών δημιουργίας δεδομένων.

Τα νευρωνικά δίκτυα αποτελούν τη φυσική επέκταση των γραμμικών μοντέλων, τα οποία μπορούν να εξαχθούν από τα νευρωνικά δίκτυα, όταν δεν υπάρχει κρυφό στρώμα και η συνάρτηση μεταφοράς της μονάδας εξόδου είναι γραμμική.

Στη περίπτωση της γραμμικής παλινδρόμησης, η επιλογή μοντέλου ανάγεται στην σωστή εκτίμηση των συντελεστών (δηλαδή των παραμέτρων) της γραμμής της παλινδρόμησης. Δεδομένων των συνηθισμένων υποθέσεων για την διακύμανση δειγματοληψίας, η καθιερωμένη στατιστική θεωρία μας επιτρέπει να ελέγξουμε την σπουδαιότητα κάθε παραμέτρου. Ακόμη και όταν εισάγεται μία δυναμική δομή, όπως είναι η περίπτωση των γραμμικών μοντέλων χρονολογικών σειρών των Box & Jenkins (1978), έχουμε διαθέσιμες επαρκώς προσδιορισμένες στατιστικές για την επιλογή μοντέλου.

Στη περίπτωση των νευρωνικών δικτύων όμως, οι συνηθισμένες υποθέσεις σχετικά με τη γραμμικότητα και τις κατανομές δειγματοληψίας δεν γίνονται, με αποτέλεσμα να καθίσταται εξαιρετικά δύσκολη κάθε προσπάθεια δημιουργίας μιας εργαλειοθήκης διαδικασιών ελέγχου για την επιλογή μοντέλου. Οι πρώιμες μέθοδοι επι-

λογής υποδείγματος (που ακόμη και σήμερα αυτές και οι παραλλαγές τους χρησιμοποιούνται ευρέως) για τον έλεγχο των «*άσχετων συνδέσεων*<sup>1</sup>», όπως το «*κλάδεμα*<sup>2</sup>», η «*κανονικοποίηση*<sup>3</sup>» και το «*πρόωρο σταμάτημα της εκπαίδευσης*», παρουσιάζουν κάποιο βαθμό υποκειμενικότητας καθώς δεν εδράζονται στην κατανομή δειγματοληψίας των παραμέτρων του δικτύου και ως εκ τούτου δεν τις εξετάζουμε εδώ περαιτέρω. Μία επισκόπηση αυτών των προσεγγίσεων δίνεται από τον Reed (1993).

Το πρώτο βήμα προς τη κατεύθυνση της δημιουργίας ενός ολοκληρωμένου και συνεκτικού πλαισίου επιλογής νευρωνικών μοντέλων έγινε από τον Moody (1992), ο οποίος περιέγραψε μία βηματική διαδικασία επιλογής του αριθμού των κρυφών μονάδων στη διαμόρφωση αναφοράς (όπως αυτή ορίστηκε στο Κεφάλαιο 4), η οποία στηριζόταν στην εκτίμηση του *κινδύνου πρόβλεψης*<sup>4</sup>, δηλαδή της τιμής της συνάρτησης μέσου τετραγωνικού σφάλματος για δεδομένα που δεν έχουν χρησιμοποιηθεί για την εκπαίδευση του δικτύου. Ο Moody περιέγραψε μία μη γραμμική επέκταση του υπολογισμού του κινδύνου πρόβλεψης, που ονόμασε γενικευμένο *σφάλμα πρόβλεψης*<sup>5</sup> (GPE).

Η εκτίμηση του κινδύνου πρόβλεψης, εκφράζει την *ικανότητα γενίκευσης* του δικτύου και παραδοσιακά χρησιμοποιείται στα πλαίσια της επιλογής των επεξηγηματικών μεταβλητών των γραμμικών μοντέλων. Πέρα από την απ' ευθείας εκτίμηση του κινδύνου πρόβλεψης, για την επιλογή του *γραμμικού* μοντέλου με την καλύτερη ικανότητα πρόβλεψης, χρησιμοποιείται μια σειρά από *κριτήρια πληροφόρησης*<sup>6</sup>, όπως τα  $C_p$  (Mallows, 1973),  $GCV$ <sup>7</sup> (Craven & Wahba, 1979; Golub *et al*, 1979),  $FPE$ <sup>8</sup> (Akaike, 1970),  $AIC$ <sup>9</sup> (Akaike, 1973),  $PSE$ <sup>10</sup> (Barron, 1984), κτλ. Όπως και στη περίπτωση της απευθείας εκτίμησης του κινδύνου πρόβλεψης τα κριτήρια πληροφόρησης είναι κατάλληλα για την επιλογή μοντέλου μεταξύ μιας σειράς ιεραρχικών μοντέλων, όπου το ένα περικλείεται στο άλλο ως ένα υπο-μοντέλο χαμηλότερης διάστασης. Το 1994 παρουσιάστηκε το κριτήριο  $NIC$ <sup>11</sup> (Murata *et al*, 1994) το ο-

<sup>1</sup> Irrelevant connections.

<sup>2</sup> Pruning.

<sup>3</sup> Regularization.

<sup>4</sup> Prediction risk.

<sup>5</sup> Generalized Prediction Error (GPE).

<sup>6</sup> Information Criteria.

<sup>7</sup> Generalized Cross Validation.

<sup>8</sup> Final Prediction Error.

<sup>9</sup> Information Criterion A.

<sup>10</sup> Predicted Squared Error.

<sup>11</sup> Network Information Criterion.

ποίο εκτιμά τον κίνδυνο πρόβλεψης μη γραμμικών μοντέλων και αποτελεί γενίκευση του κριτηρίου GPE του Moody<sup>12</sup>.

Μία άλλη προσέγγιση στο πρόβλημα της επιλογής νευρωνικών μοντέλων είναι ο στατιστικός έλεγχος της υπόθεσης ότι συγκεκριμένες συνδέσεις του δικτύου είναι μηδενικές. Επιλέγοντας τον κατάλληλο συνδυασμό συνδέσεων μπορούμε ουσιαστικά με αυτή τη προσέγγιση να ελέγξουμε την υπόθεση ότι κάποιες κρυφές μονάδες ή κάποιες μονάδες εισόδου είναι *πλεονάζουσες*<sup>13</sup> και συνεπώς μπορούν να αφαιρεθούν. Καθώς τα νευρωνικά δίκτυα είναι μη γραμμικά μοντέλα παλινδρόμησης, μπορούν κατ' αρχήν να εφαρμοστούν οι τυπικές διαδικασίες ελέγχου της σημαντικότητας των παραμέτρων, όπως οι έλεγχοι Wald και πολλαπλασιαστών Lagrange (LM). Δύο χαρακτηριστικοί έλεγχοι αυτού του είδους είναι ο έλεγχος που προτάθηκε από τον White (1989a) και ο έλεγχος μη γραμμικότητας TLG που προτάθηκε από τους Terasvirta, Lin και Granger (1993). Για να εφαρμοστούν όμως οι διαδικασίες ελέγχου Wald, είναι απαραίτητο να γνωρίζουμε την ασύμπτωτη κατανομή των παραμέτρων του δικτύου. Αυτή η κατανομή εξήχθη από τον White (1989b) και αποτελεί τη βάση του υπολογισμού και του κριτηρίου NIC.

Εδώ θα πρέπει να αναφέρουμε ότι η χρήση τόσο των παραπάνω ελέγχων όσο και των κριτηρίων πληροφόρησης και του κινδύνου πρόβλεψης δεν δικαιολογείται θεωρητικά όταν υπάρχουν πλεονάζουσες κρυφές μονάδες, έτσι λοιπόν θα πρέπει υποχρεωτικά να διενεργούνται πρώτα οι έλεγχοι κρυφών μονάδων πριν από τους ελέγχους των συνδέσεων εισόδου. Η σωστή επιλογή του αριθμού των κρυφών μονάδων διασφαλίζει ότι τελικά το μοντέλο θα συλλάβει όλες τις υφιστάμενες μη γραμμικότητες, χωρίς τη χρήση *άσχετων* κρυφών μονάδων. Από την άλλη η απαλοιφή μη απαραίτητων συνδέσεων εισόδου είναι ένας τρόπος ελέγχου της περιπλοκότητας του μοντέλου και εξαίρεσης των *άσχετων* εισόδων. Δυστυχώς μέχρι στιγμής δεν έχει προταθεί κάποιο τεστ για τον έλεγχο και των δύο ταυτόχρονα.

Εναλλακτικά η εκτίμηση του κινδύνου πρόβλεψης μπορεί να γίνει χρησιμοποιώντας *στατιστικές μεθοδολογίες επαναλαμβανόμενης δειγματοληψίας*<sup>14</sup>, όπως η μέθοδος της *σταυροειδούς επαλήθευσης*<sup>15</sup> (CV) (Stone, 1974) και η μέθοδος της *επαναλαμβανόμενης δειγματοληψίας με επανατοποθέτηση*<sup>16</sup> (Efron, 1981).

Οι Zapranis και Refenes επεκτείνοντας το υπόβαθρο που έθεσε ο Moody το 1992 (Refenes, Zapranis, 1999; Zapranis & Refenes, 1999), πρότειναν ένα ολοκληρωμένο πλαίσιο για την επιλογή νευρωνικού υποδείγματος, το οποίο ενσωματώνει τα ακόλουθα στάδια:

---

<sup>12</sup> Όταν ο θόρυβος είναι προσθετικός το NIC ταυτίζεται με το GPE.

<sup>13</sup> Ο όρος συναντάται στην Αγγλόφωνη βιβλιογραφία ως irrelevant. Στο βιβλίο αυτό αναφερόμαστε σ' αυτού του είδους τις συνδέσεις και μονάδες ως «πλεονάζουσες» ή «άσχετες».

<sup>14</sup> Resampling schemes.

<sup>15</sup> Cross-validation.

<sup>16</sup> Bootstrap.

- επιλογή του αρχικού υποδείγματος από μία ακολουθία μοντέλων αυξανόμενης περιπλοκότητας,
- επιλογή ερμηνευτικών μεταβλητών μέσω της εκτίμησης της στατιστικής σημαντικότητάς τους,
- έλεγχος της επάρκειας του τελικού υποδείγματος.

Κατά το πρώτο στάδιο επιλέγεται ο βέλτιστος αριθμός κρυφών μονάδων, για ένα πλήρως συνδεδεμένο δίκτυο με τη διαμόρφωση αναφοράς, στη βάση της ελαχιστοποίησης του κινδύνου πρόβλεψης. Η διαδικασία αρχίζει με ένα πλήρως συνδεδεμένο δίκτυο με 0 κρυφές μονάδες (γραμμικό μοντέλο), και έπειτα προστίθενται επιπλέον κρυφές μονάδες, για όσο ο κίνδυνος πρόβλεψης εξακολουθεί να μειώνεται (Εικόνα 1). Ο αριθμός των κρυφών μονάδων που αντιστοιχεί στην ελάχιστη τιμή του κινδύνου πρόβλεψης είναι αυτός που επιλέγεται τελικά.

Έπειτα, θα πρέπει να αφαιρεθούν οι πλεονάζουσες συνδέσεις εισόδου ώστε το μοντέλο να καταστεί *τοπικά ταυτοποιημένο*<sup>17</sup>. Ονομάζουμε έτσι ένα δίκτυο που δεν περιέχει πλεονάζουσες συνδέσεις και άρα αντιστοιχεί σε μία «μοναδική» λύση<sup>18</sup> και όχι σε ένα «πλάτωμα» ισοδύναμων λύσεων. Ο λόγος είναι ότι η κατανομή των παραμέτρων των δικτύων που περιέχουν πλεονάζουσες συνδέσεις δεν είναι ασύμπτωτα κανονική όπως απέδειξε ο White (1989b), αλλά ανήκει στην οικογένεια των κατανομών «Mixed Gaussian» των *μερικώς ταυτοποιημένων μοντέλων*<sup>19</sup> (Phillips, 1989). Η επίπτωση αυτού του γεγονότος είναι ότι η χρήση της αλγεβρικής εκτίμησης του κινδύνου πρόβλεψης ή των κριτηρίων πληροφόρησης για την επιλογή μοντέλου δεν θα μας δώσει απαραίτητα το μοντέλο με την καλύτερη ικανότητα γενίκευσης. Οι πλεονάζουσες συνδέσεις εισόδου μπορούν να απομακρυνθούν χρησιμοποιώντας κάποια από τις μεθοδολογίες κλαδέματος, όπως για παράδειγμα τον αλγόριθμο ICE (Zarpanis & Haramis, 2001). Εναλλακτικά, οι συνδέσεις εισόδου μπορούν να αφαιρούνται μία προς μία έως φθάσουμε σε κάποια ελάχιστη τιμή του κινδύνου πρόβλεψης. Είναι προφανές ότι η δεύτερη προσέγγιση είναι σημαντικά πιο χρονοβόρα.

Οι τεχνικές επαναλαμβανόμενης δειγματοληψίας δεν βασίζονται σε υποθέσεις κατανομών και δεν επηρεάζονται από προβλήματα ταυτοποίησης. Από την άλλη, είναι υπολογιστικά ακριβές μέθοδοι εάν υλοποιηθούν εξαντλητικά για κάθε πιθανό

<sup>17</sup> Locally identified. Ο όρος «τοπικά» (locally) υπονοεί την πιθανότητα η λύση να αντιστοιχεί σε ένα τοπικό ελάχιστο και όχι στο ολικό ελάχιστο. Το γεγονός αυτό δεν επηρεάζει την ασύμπτωτη κατανομή των παραμέτρων του δικτύου που δίνει ο White (1989b).

<sup>18</sup> Στη πράξη καμία λύση δεν μπορεί να θεωρηθεί «μοναδική», καθώς εξαιτίας της συμμετρίας της αρχιτεκτονικής του δικτύου μπορούν δύο κρυφές μονάδες να ανταλλάξουν τη θέση τους, χωρίς να επηρεαστεί η απόκριση του δικτύου, ενώ θα υπάρξει αντιμετάθεση των στοιχείων του διανύσματος των παραμέτρων, μια νέα λύση δηλαδή. Από την άλλη, αυτή η αιτία πρόκλησης μη-ταυτοποίησης δεν είναι ανησυχητική καθώς δεν επηρεάζει την ασύμπτωτη κατανομή των παραμέτρων που δίνει ο White (1989b).

<sup>19</sup> Partially identified models.



συνδυασμό κρυφών μονάδων και συνδέσεων εισόδου. Σε προβλήματα με υψηλό αριθμό επεξηγηματικών μεταβλητών, για να γίνει εφικτή η χρήση αυτών των τεχνικών, χρησιμοποιούνται στατιστικές τεχνικές, όπως η ανάλυση κυρίων συνιστωσών καθώς και βηματικές τεχνικές επιλογής μεταβλητών γραμμικής παλινδρόμησης, για την μείωση του αρχικού αριθμού των επεξηγηματικών μεταβλητών. Στη συνέχεια γίνεται πλήρης έρευνα πάνω στον (ελαττωμένο τώρα) χώρο των μη γραμμικών μοντέλων χρησιμοποιώντας τεχνικές επαναλαμβανόμενης δειγματοληψίας.

Κατά το δεύτερο στάδιο α) γίνεται ο προσδιορισμός της καταλληλότητας της μεταβλητής εισόδου για το υπόδειγμα, β) η εκτίμηση της δειγματικής διακύμανσης της μέτρησης καταλληλότητας και γ) ο έλεγχος της υπόθεσης ότι η μεταβλητή είναι ακατάλληλη. Για τον προσδιορισμό της καταλληλότητας της μεταβλητής εισόδου οι (Refenes, Zapranis, 1999; Zapranis & Refenes, 1999) εξέτασαν τη χρήση μιας σειράς κριτηρίων ευαισθησίας<sup>20</sup> και κριτηρίων ευαισθησίας καταλληλότητας μοντέλου<sup>21</sup>. Η εκτίμησης της δειγματικής διακύμανσης μπορεί να γίνει είτε χρησιμοποιώντας τεχνικές επαναλαμβανόμενης δειγματοληψίας είτε με τυχαία δειγματοληψία από την ασύμπτωτη κατανομή των παραμέτρων. Με δεδομένη την δειγματική διακύμανση μπορούμε να αξιολογήσουμε την σημαντικότητα των μετρήσεων καταλληλότητας.

Η επιλογή των ερμηνευτικών μεταβλητών θα μπορούσε να φυσικά να στηριχθεί επάνω σε οποιαδήποτε από τις προσεγγίσεις που προαναφέραμε, δηλαδή α) κριτήρια πληροφόρησης, β) εκτίμηση του κινδύνου πρόβλεψης αλγεβρικά ή με μεθόδους επαναλαμβανόμενης δειγματοληψίας γ) έλεγχος υποθέσεων. Οι Anders και Korn (1999) συνέκριναν την επίδοση όλων παραπάνω προσεγγίσεων σε τεχνητά δεδομένα. Αναφέρουν, ότι ο έλεγχος μη γραμμικότητας TLG γενικά δίνει την σταθερότερη συμπεριφορά του επιλεγμένου δικτύου, ενώ ο έλεγχος του White την χειρότερη συμπεριφορά. Τα κριτήρια πληροφόρησης και η σταυροειδής επαλήθευση δίνουν πολύ ακριβή αποτελέσματα για μερικά μοντέλα, ενώ τείνουν να υπερπροσαρμόζουν ή να υποπροσαρμόζουν σε άλλα. Οι (Refenes, Zapranis, 1999; Zapranis & Refenes, 1999) αναφέρουν ότι το bootstrap τείνει να υπερεκτιμά την διακύμανση δείγματος, ως αποτέλεσμα της ύπαρξης τοπικών ελαχίστων.

Κατά το τρίτο στάδιο ελέγχεται εάν ευσταθούν οι αρχικές υποθέσεις σχετικά με την πραγματική υποκείμενη σχέση που δημιούργησε τις διαθέσιμες παρατηρήσεις, χωρίς όμως η ικανοποίηση των ελέγχων καταλληλότητας του υποδείγματος να αποτελεί επαρκή συνθήκη για τον χαρακτηρισμό ενός υποδείγματος ως βέλτιστου.

Στις επόμενες ενότητες εστιάζομαστε κυρίως στην εκτίμηση του κινδύνου πρόβλεψης αλγεβρικά και με μεθόδους επαναλαμβανόμενης δειγματοληψίας, που όπως είδαμε μπορούν να χρησιμοποιηθούν και για τα δύο στάδια της διαδικασίας επιλο-

---

<sup>20</sup> Sensitivity criteria.

<sup>21</sup> Model-fitness sensitivity criteria.

γής νευρωνικού υποδείγματος, καθώς και στους ελέγχους καταλληλότητας του υποδείγματος. Για τα διάφορα κριτήρια ευαισθησίας και κριτήρια ευαισθησίας καταλληλότητας μοντέλου που μπορούν να χρησιμοποιηθούν εναλλακτικά, ο αναγνώστης μπορεί να ανατρέξει στους (Refenes, Zapranis, 1999; Zapranis & Refenes, 1999).

## 5.2 Ο Κίνδυνος Πρόβλεψης

Ο κίνδυνος πρόβλεψης, του δικτύου  $g_\lambda(\mathbf{x}; \hat{\mathbf{w}}_n)$ , που θα συμβολίζουμε ως  $P_\lambda$ , είναι η αναμενόμενη απόδοσή του δικτύου σε μελλοντικά δεδομένα και δίνεται από την παρακάτω έκφραση:

$$P_\lambda = \int [\varphi(\mathbf{x}) - g_\lambda(\mathbf{x}; \hat{\mathbf{w}}_n)]^2 p(\mathbf{x}) d\mathbf{x} + \sigma_\varepsilon^2 \quad (1)$$

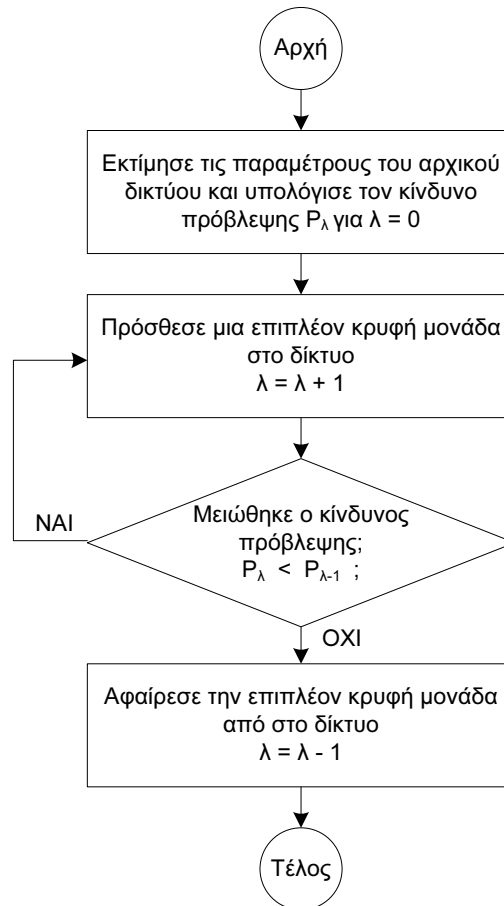
όπου  $\sigma_\varepsilon^2$  είναι η διακύμανση των ασυσχέτιστων τυχαίων σφαλμάτων με μηδενική μέση τιμή. Η προσέγγιση περιορισμένου δείγματος (finite sample) του κινδύνου πρόβλεψης είναι:

$$P_\lambda \approx E \left[ \frac{1}{n} \sum_{i=1}^n (y_i^* - g_\lambda(\mathbf{x}_i^*; \hat{\mathbf{w}}_n))^2 \right] \quad (2)$$

όπου  $(y_i^*, \mathbf{x}_i^*)$  είναι νέες παρατηρήσεις που δεν έχουν χρησιμοποιηθεί για την κατασκευή του μοντέλου  $g_\lambda(\mathbf{x}; \hat{\mathbf{w}}_n)$ . Σύμφωνα με τα παραπάνω ο κίνδυνος πρόβλεψης αποτελεί ουσιαστικά μία εκτίμηση της ικανότητας γενίκευσης του δικτύου.

Οι εκτιμήσεις του κινδύνου πρόβλεψης διαφορετικών νευρωνικών μοντέλων μπορούν να χρησιμοποιηθούν για την επιλογή του βέλτιστου αριθμού κρυφών μονάδων  $\lambda$  στα πλαίσια μιας βηματικής διαδικασίας που πραγματοποιεί μια κατευθυνόμενη αναζήτηση, καθώς μια διεξοδική αναζήτηση δεν είναι εφικτή. Έστω λοιπόν μια αυξητική δομημένη ιεραρχία ένθετων κλάσεων δικτύου  $S_1 \subset S_2 \subset \dots \subset S_k \dots$ .

Καθώς δεν απαιτείται πάνω από ένα κρυμμένο στρώμα ώστε το νευρωνικό δίκτυο να έχει τη δυνατότητα να εκτιμήσει οποιαδήποτε συνάρτηση, κάθε κλάση περιλαμβάνει όλα τα πιθανά δίκτυα ενός στρώματος με την ίδια αρχιτεκτονική  $A_\lambda$ . Ο αριθμός των κρυμμένων μονάδων  $\lambda$  χαρακτηρίζει με τρόπο μοναδικό κάθε κλάση. Με την προϋπόθεση ότι για κάθε αρχιτεκτονική  $A_\lambda$  το δίκτυο θα πρέπει να εκπαιδευτεί μέχρι να συγκλίνει σε κάποια λύση, πραγματοποιούμε μια περιορισμένη αναζήτηση στον συναρτησιακό χώρο που καλύπτει η αρχιτεκτονική του δικτύου.



**Εικόνα 1:** Επιλογή αριθμού κρυφών μονάδων στη βάση της αρχής του ελάχιστου κινδύνου πρόβλεψης.

Αυτό γίνεται επειδή από κάθε κλάση  $\mathcal{S}_\lambda$  επιλέγουμε μόνο ένα μέλος με διάνυσμα παραμέτρων που εκτιμάται από την σχέση:

$$\hat{\mathbf{w}}_n = \arg \min \{L_n(\mathbf{w}) : \mathbf{w} \in \mathbf{W}\}, \quad \mathbf{W} \subseteq \mathcal{R}^p, p = h(A_\lambda) \quad (3)$$

όπου  $L_n(\mathbf{w})$  είναι μια γενική *συνάρτηση απώλειας*<sup>22</sup>, που ελαχιστοποιείται κατά την διάρκεια της εκπαίδευσης του δικτύου, που ενδεχομένως να περιλαμβάνει και όρους *κανονικοποίησης*<sup>23</sup> όπως η *αποσύνθεση βαρών*<sup>24</sup>. Σ' αυτή την περίπτωση η συνάρτηση που ελαχιστοποιείται έχει την μορφή:

$$L_n(\mathbf{w}) = \frac{1}{2n} \sum_{i=1}^n (y_i - g_\lambda(\mathbf{x}_i; \mathbf{w}))^2 + a \sum_{i=1}^p w_i \quad (4)$$

όπου  $a > 0$  είναι η *παράμετρος κανονικοποίησης*<sup>25</sup>. Στην ανάλυση που ακολουθεί η συνάρτηση απώλειας που χρησιμοποιεί η σχέση (3) ταυτίζεται με την συνάρτηση του μέσου τετραγώνου σφάλματος, που είδαμε στο προηγούμενο κεφάλαιο. Δηλαδή:

$$L_n(\mathbf{w}) \equiv E_n(\mathbf{w}) = \frac{1}{2n} \sum_{i=1}^n (y_i - g_\lambda(\mathbf{x}_i; \mathbf{w}))^2 \quad (5)$$

Ξεκινώντας από  $\lambda = 0$  και προσθέτοντας μία κρυφή μονάδα κάθε φορά, όπως στην περιγράφεται στην Εικόνα 1, επιλέγουμε τον αριθμό κρυφών μονάδων που ελαχιστοποιεί τον κίνδυνο πρόβλεψης  $P_\lambda$ . Στις επόμενες δύο ενότητες εξετάζουμε τις δύο βασικές προσεγγίσεις εκτίμησης του κινδύνου πρόβλεψης: α) την αλγεβρική εκτίμηση και β) την εκτίμηση με μεθόδους επαναλαμβανόμενης δειγματοληψίας.

### 5.3 Αλγεβρική Εκτίμηση του Κινδύνου Πρόβλεψης

Εάν γνωρίζουμε τη κατανομή πυκνότητας πιθανότητας των παραμέτρων του δικτύου (των συνδέσεων του) είναι δυνατόν να υπολογίσουμε μία αλγεβρική εκτίμηση του κινδύνου πρόβλεψης. Την κατανομή αυτή την εξήγαγε ο White (1989b). Ο White απέδειξε ότι, εάν υπάρχει ένα *μοναδικό* διάνυσμα παραμέτρων  $\mathbf{w}_0$ , το οποίο οδηγεί στη βέλτιστη προσέγγιση της συνάρτησης  $\varphi(\mathbf{x})$  από ένα συγκεκριμένο νευρωνικό μοντέλο  $g_\lambda(\mathbf{x}; \mathbf{w})$ , τότε το διάνυσμα  $\hat{\mathbf{w}}_n$  που εκτιμάται ελαχιστοποιώντας την (5) αποτελεί ένα συνεπή εκτιμητή του διανύσματος  $\mathbf{w}_0$ . Επιπλέον, το διάνυσμα  $\hat{\mathbf{w}}_n$  ακολουθεί ασύμπτωτα την κανονική κατανομή με μέση τιμή το διάνυσμα  $\mathbf{w}_0$  και μητρώο διακύμανσης  $(1/n)\mathbf{C}$ , ή διαφορετικά διατυπωμένο:

<sup>22</sup> Loss function.

<sup>23</sup> Απόδοση στα ελληνικά του όρου regularization.

<sup>24</sup> Απόδοση στα ελληνικά του όρου weight decay.

<sup>25</sup> Regularization parameter.

$$\sqrt{n}(\hat{\mathbf{w}}_n - \mathbf{w}_0) \sim N(0, \mathbf{C}) \quad (6)$$

όπου

$$\mathbf{C} = \mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-1} \quad (7)$$

$$\mathbf{A} = E[\nabla^2 L(\mathbf{w}_0)] \quad (8)$$

$$\mathbf{B} = E[\nabla L(\mathbf{w}_0) \nabla L(\mathbf{w}_0)^T] \quad (9)$$

Τα σύμβολα  $\nabla$  και  $\nabla^2$  στις παραπάνω σχέσεις δηλώνουν αντίστοιχα το βαθμωτό ανάδελτα ( $p \times 1$ ) και τον εσσιανό συντελεστή ( $p \times p$ ) σε σχέση με το  $\mathbf{w}_0$  (όπου  $p$  είναι ο αριθμός των παραμέτρων του δικτύου). Αναλυτικές εκφράσεις για τις παραγώγους των απαιτούνται από τις σχέσεις (8) και (9) δίνονται από τους (Zapranis & Refenes, 1999).

Η πιο συνηθισμένη προσέγγιση για την αλγεβρική εκτίμηση του κινδύνου πρόβλεψης συνίσταται στον υπολογισμό των δύο πρώτων όρων της επέκτασης κατά Taylor της εμπειρικής συνάρτησης σφάλματος  $L(\hat{\mathbf{w}}_n)$ , γύρω από το διάνυσμα  $\mathbf{w}_0$  και στη συνέχεια στον υπολογισμό της προσδοκώμενης τιμής της. Μετά από μερικούς αριθμητικούς μετασχηματισμούς (Murata *et al*, 1994; Zapranis & Refenes, 1999) εξάγεται η παρακάτω σχέση της αναλυτικής εκτίμησης για τον κίνδυνο πρόβλεψης:

$$E[L(\hat{\mathbf{w}}_n)] \approx E[L_n(\hat{\mathbf{w}}_n)] + \frac{tr \mathbf{A}^{-1} \mathbf{B}}{n} \quad (10)$$

Το ίχνος  $tr \mathbf{A}^{-1} \mathbf{B}$  αντιστοιχεί στη διαφορά μεταξύ του κινδύνου πρόβλεψης και του προσδοκώμενου σφάλματος εκπαίδευσης. Όταν το  $tr \mathbf{A}^{-1} \mathbf{B}$  είναι αρκετά μεγάλο σε σχέση με το προσδοκώμενο σφάλμα εκπαίδευσης, τότε έχουμε υπερπροσαρμογή στα δεδομένα. Στην περίπτωση αυτή ο αριθμός των παραμέτρων του δικτύου είναι μεγαλύτερος από αυτόν που απαιτείται, δεδομένων των διαθέσιμων παρατηρήσεων. Η σχέση (10) συμπυκνώνει την ουσία της επιλογής υποδείγματος βασισμένη στον κίνδυνο πρόβλεψης. Τα πολύπλοκα υποδείγματα είναι σχετικά αμερόληπτα (δηλαδή το σφάλμα εκπαίδευσης είναι χαμηλό), αλλά υποφέρουν από υψηλή μεταβλητότητα (ο όρος  $tr \mathbf{A}^{-1} \mathbf{B}$  είναι μεγάλος). Ο κίνδυνος πρόβλεψης ως το άθροισμα

αυτών των δύο ποσοτήτων προσπαθεί να πετύχει μια ισορροπία μεταξύ αυτών των δύο αντίθετων δυνάμεων.

Στην πράξη, το  $tr\mathbf{A}^{-1}\mathbf{B}$  είναι άγνωστο. Τα μητρώα  $\mathbf{A}$  και  $\mathbf{B}$  θα πρέπει να αντικατασταθούν με τους ασθενώς συνεπίεις εκτιμητές τους  $\hat{\mathbf{A}}_n$  και  $\hat{\mathbf{B}}_n$ , που υπολογίζονται για τα δεδομένα  $D_n = \{(x_i, y_i)\}_{i=1}^n$  και τον εκτιμητή ελάχιστης ασυνέπειας  $\hat{\mathbf{w}}_n$ . Το επακόλουθο της αντικατάστασης των  $\mathbf{A}$  και  $\mathbf{B}$  με τους εκτιμητές τους  $\hat{\mathbf{A}}_n$  και  $\hat{\mathbf{B}}_n$ , είναι ότι χρειάζεται μεγάλος αριθμός δεδομένων για να επιτευχθεί μια καλή προσέγγιση του  $tr\mathbf{A}^{-1}\mathbf{B}$ . Η προσδοκία  $E[L(\hat{\mathbf{w}}_n)]$  είναι επίσης άγνωστη και στην πράξη αντικαθίσταται από το σφάλμα εκπαίδευσης  $L_n(\hat{\mathbf{w}}_n)$ . Έτσι, η εκτίμησή μας για τον κινδύνου πρόβλεψης και την συσχετιζόμενη διακύμανση γίνεται

$$\hat{P}_\lambda \equiv \hat{E}[L(\hat{\mathbf{w}}_n)] \approx L_n(\hat{\mathbf{w}}_n) + \frac{tr\hat{\mathbf{A}}_n^{-1}\hat{\mathbf{B}}_n}{n} \quad (11)$$

$$\hat{\text{var}}[L(\hat{\mathbf{w}}_n)] \approx \frac{(tr\hat{\mathbf{A}}_n^{-1}\hat{\mathbf{B}}_n)^2}{2n^2} \quad (12)$$

όπου τα μητρώα  $\hat{\mathbf{A}}_n$  και  $\hat{\mathbf{B}}_n$ , δίνονται από τις σχέσεις (13) και (14).

$$\hat{\mathbf{A}}_n = \sum_{i=1}^n \nabla^2 L_n(\hat{\mathbf{w}}_n) \quad (13)$$

$$\hat{\mathbf{B}}_n = \sum_{i=1}^n \nabla L_n(\hat{\mathbf{w}}_n) \nabla L_n(\hat{\mathbf{w}}_n)^T \quad (14)$$

Η σχέση (11) αποτελεί την βάση για την αλγεβρική εκτίμηση του κινδύνου πρόβλεψης. Για μια παραπλήσια σχέση δεξ επίσης το GPE του Moody (1992). Η χρήση της σχέσης (11) προϋποθέτει ότι το διάνυσμα  $\hat{\mathbf{w}}_n$  αποτελεί μοναδική λύση. Δεν αντιστοιχεί απαραίτητα σε ολικό ελάχιστο, αλλά είναι απαραίτητο να μην περιλαμβάνει πλεονάζουσες συνδέσεις. Μπορούμε να διακρίνουμε λοιπόν δύο στάδια κατά τον προσδιορισμό της τοπολογίας του δικτύου στη βάση της αλγεβρικής εκτίμησης του κινδύνου πρόβλεψης από τη σχέση (11): *i*) προσδιορισμός του αριθμού των κρυφών μονάδων  $\lambda$  και *ii*) αφαίρεση των άσχετων συνδέσεων του δικτύου.