ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ - ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ & ΤΕΧΝΟΛΟΓΙΑΣ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ & ΤΕΧΝΟΛΟΓΙΑΣ ΤΗΣ ΠΛΗΡΟΦΟΡΙΑΣ
Ακαδημαϊκό Ετος 2015 – Εξεταστική περίοδος Ιουνίου

Αναγνώριση προτύπων Ι

Θέμα 1 (4 μονάδες): Εστω ότι διατίθενται τα δισδιάστατα παραδείγματα εκπαίδευσης τριών κατηγοριών: Ω1={(0,0),(1,1),(2,2)}, Ω2={(0,1),(1,2),(2,3)} και Ω3={(1,0),(2,1),(3,2)}. Η συνάρτηση απόστασης που εκτιμά την απόσταση του διανύσματος **x** από το **y** δίνεται από την σχέση

$$d(\mathbf{x},\mathbf{y})=|\mathbf{xy}|^2+|\mathbf{x}-\mathbf{y}|^2$$

όπου **xy** είναι το εσωτερικό γινόμενο των δύο διανυσμάτων.
(α. 2 μοναδες) Υπολογίστε ένα εικονικό κέντρο ανά κατηγορία προτύπων
(β. 2 μονάδες) Υπολογίστε το ελάχιστο σφάλμα του ταξινομητή ελάχιστης απόστασης.

Λύση θέματος 1:
Το ερώτημα (α) ζητάει να βρεθούν ένα εικονικό κέντρο ανα κατηγορία. Συνεπώς πρέπει πρώτα να βρεθεί η σχεση που συνδέει το κέντρο από την αριθμητική τιμή των παραδειγμάτων. Το κριτήριο εκτίμησης του εικονικού κέντρου είναι η εύρεση διανύσματος που ελαχιστοποιεί την αθροιστική απόστασή του από τα παραδείγματα.
Συνεπώς, αν $\mathbf{y}=(y_1,y_2)$ είναι το εικονικό διάνυσμα και $\mathbf{x}_i$, i =1,N τα παραδείγματα εκαπαίδευσης τότε βάσει του κριτηρίου έχω:

$$\frac{\partial}{\partial y_j}\sum_i d(x_i,y)=0\Rightarrow\sum_i\frac{\partial}{\partial y_j}(|x_i y|^2+|x_i-y|^2)=\sum_i(\frac{\partial}{\partial y_j}|x_i y|^2+\frac{\partial}{\partial y_j}|x_i-y|^2)=0\Rightarrow$$

Γιά κάθε άγνωστη παράμετρο ξεχωριστά y_1 και y_2 έχω:

$$\sum_i(\frac{\partial}{\partial y_1}(x_{1i}y_1+x_{2i}y_2)^2+\frac{\partial}{\partial y_1}((x_{1i}-y_1)^2+(x_{2i}-y_2)^2))=0\Rightarrow$$

$$2\sum_i((x_{1i}y_1+x_{2i}y_2)x_{1i}+(y_1-x_{1i}))=0\Rightarrow\sum_i x_{1i}^2y_1+\sum_i x_{1i}x_{2i}y_2+N\,y_1-\sum_i x_{1i}=0\Rightarrow$$

$$(\sum_i x_{1i}^2+N)\,y_1+(\sum_i x_{1i}x_{2i})\,y_2=\sum_i x_{1i} \qquad (1)$$

Με τον ίδιο τρόπο προκύπτει η συμμετρική σχέση παραγωγίζοντας την συνάρτηση ως προς y_2:

$$(\sum_i x_{1i}x_{2i})\,y_1+(\sum_i x_{2i}^2+N)\,y_2=\sum_i x_{2i} \qquad (2)$$

Για κάθε κατηγορία ξεχωριστά υπολογιζεται και επιλύεται το σύστημα των γραμμικών εξισωσεων (1) και (2)

Κατηγορία Ω_1: $(5+3)\,y_1+5\,y_2=3$, $5\,y_1+(5+3)\,y_2=3$ Συνεπώς $y_1 = y_2 = 3/13$
Κατηγορία Ω_2: $(5+3)\,y_1+8\,y_2=3$, $8\,y_1+(14+3)\,y_2=6$ Συνεπώς $y_1 = 1/24$, $y_2 = 1/3$
Κατηγορία Ω_3: $(14+3)\,y_1+8\,y_2=6$, $8\,y_1+(5+3)\,y_2=5$ Συνεπώς $y_1 = 1/9$, $y_2 = 37/72$

Τα εικονικά κέντρα των τριών κατηγοριών είναι:

Κατηγορία Ω_1:	0.2307692308	0.2307692308
Κατηγορία Ω_2:	0.0416666667	0.3333333333
Κατηγορία Ω_3:	0.1111111111	0.5138888889

Το ερώτημα (β) ζητάει να υπολογίσουμε την απόσταση όλων των παραδειγμάτων από τα εικονικά κέντρα, να βρούμε το κοντινότερο κέντρο και να υπολογίσουμε το σφάλμα ταξινόμησης ταξινομώντας κάθε παράδειγμα στον πλησιέστερο εικονικό κέντρο.

Κατ. Ω1			d(x,Ω1)	d(x,Ω2)	d(x,Ω3)	Ταξινόμηση	Αξιολόγη
x1	0	0	0.1065	0.1128	0.2764	Ω1	Σωστό
x2	1	1	1.3964	1.5035	1.4171	Ω1	Σωστό
x3	2	2	7.1124	7.1753	7.3389	Ω1	Σωστό
Κατ. Ω2							
x1	0	1	0.6982	0.5573	0.5127	Ω3	Λάθος
x2	1	2	4.2012	4.1979	4.2957	Ω2	Σωστό
x3	2	3	12.1302	12.1198	12.8600	Ω2	Σωστό
Κατ. Ω3							
x1	1	0	0.6982	1.0313	1.0666	Ω1	Λάθος
x2	2	1	4.2012	4.4531	4.3461	Ω1	Λάθος
x3	3	2	12.1302	12.1563	12.4068	Ω1	Λάθος

Το ελάχιστο σφάλμα του ταξινομητή είναι 4/9 * 100% = 44.44%

Θέμα 2 (4 μονάδες): Εστω ότι διατίθενται τα δισδιάστατα παραδείγματα εκπαίδευσης των τριών κατηγοριών του θέματος 1. Γνωρίζετε ότι: η συνάρτηση πυκνότητας πιθανότητας των προτύπων κάθε κατηγορίας ακολουθεί κανονική κατανομή, οι δύο συνιστώσες του διανύσματος των προτύπων είναι στατιστικά ανεξάρτητες, τα πρότυπα της πρώτης κατηγορίας εμφανίζονται με διπλάσια συχνότητα από αυτά της δεύτερης, τα πρότυπα της δεύτερης κατηγορίας εμφανίζονται με διπλάσια συχνότητα από αυτά της τρίτης.
(α. 2 μοναδες) Να ορίσετε το κριτήριο ταξινόμησης ελάχιστου σφάλματος
(β. 2 μοναδες) Το άγνωστο πρότυπο (4,4.5) σε ποιά κατηγορία θα ταξινομηθεί με το κριτήριο ταξινόμησης ελάχιστου σφάλματος;

Λύση θέματος 2:
Το ερώτημα (α)
Εφόσον οι δεσμευμένες συναρτήσεις πυκνότητας πιθανότητας των κατηγοριών ακολουθούν όλες την κανονική κατανομή και οι συνιστώσες είναι στατιστικά αναξάρτητες τότε έχουμε την εξής απλοποίηση του κριτηρίου ταξινόμησης ελάχιστου σφάλματος:

$$\Omega_k = \underset{j}{argmax}\, p(\Omega_j|x) = \underset{j}{argmax}\, p(x|\Omega_j)\, p(\Omega_j) = \underset{j}{argmax}\, p(x_1|\Omega_j)\, p(x_2|\Omega_j)\, p(\Omega_j)$$

$$\Omega_k = \underset{j}{argmax}\, G(x_{1,} m_{1j}, s_{1j}) G(x_2, m_{2j}, s_{2j})\, p(\Omega_j)$$

Οπου G(.) είναι η Gaussian συνάρτηση πυκνότητας πιθανότητας που χαρακτηρίζεται πλήρως από την μέση τιμή και την διασπορά της.

Από την εκφώνηση γνωρίζουμε επίσης ότι οι a prioi πιθανότητες των κατηγοριών υπακούουν στους εξής περιορισμούς:

$$p(\Omega_1) = 2\, p(\Omega_2) \qquad p(\Omega_2) = 2\, p(\Omega_3) \qquad p(\Omega_1) + p(\Omega_2) + p(\Omega_3) = 1$$

Συνεπώς

$$4\, p(\Omega_3) + 2\, p(\Omega_3) + p(\Omega_3) = 1 \Rightarrow 7\, p(\Omega_3) = 1 \Rightarrow p(\Omega_3) = 1/7, p(\Omega_2) = 2/7, p(\Omega_1) = 4/7$$

Οι παράμετροι της Gaussian κατανομής υπολογίζονται απο τις διαθέσιμες μετρήσεις

Κατηγορία Ω_1

$G(x_{1,} m_{11}, s_{11}) : m_{11} = (0+1+2)/3 = 1, s_{11}^2 = ((0-1)^2 + (1-1)^2 + (2-1)^2)/(3-1) = 2/2 = 1$

$G(x_{1,} m_{12}, s_{12}) : m_{12} = (0+1+2)/3 = 1, s_{12}^2 = ((0-1)^2 + (1-1)^2 + (2-1)^2)/(3-1) = 2/2 = 1$

Κατηγορία Ω_2

$G(x_{1,} m_{21}, s_{21}) : m_{21} = (0+1+2)/3 = 1, s_{21}^2 = ((0-1)^2 + (1-1)^2 + (2-1)^2)/(3-1) = 2/2 = 1$

$G(x_{1,} m_{22}, s_{22}) : m_{22} = (1+2+3)/3 = 2, s_{22}^2 = ((1-2)^2 + (2-2)^2 + (3-2)^2)/(3-1) = 2/2 = 1$

Κατηγορία Ω_3

$G(x_{1,} m_{31}, s_{31}) : m_{31} = (1+2+3)/3 = 2, s_{31}^2 = ((1-2)^2 + (2-2)^2 + (3-2)^2)/(3-1) = 2/2 = 1$

$G(x_{1,} m_{32}, s_{32}) : m_{32} = (0+1+2)/3 = 1, s_{32}^2 = ((0-1)^2 + (1-1)^2 + (2-1)^2)/(3-1) = 2/2 = 1$

Αντικαθιστώντας τις αριθμητικές τιμές στις συναρτήσεις πυκνότητας πιθανότητας έχουμε το ακόλουθο κριτήριο ελάχιστου σφάλματος:

Αν $(x_{1,} x_2)$ είναι οι αριθμητικές τιμές του άγνωστου πρότυπου, τότε αν ισχύει

$2G(x_{1,} 1, 1) G(x_2, 1, 1) > G(x_{1,} 1, 1) G(x_2, 2, 1)$ και $4G(x_{1,} 1, 1) G(x_2, 1, 1) > G(x_{1,} 2, 1) G(x_2, 1, 1)$

το πρότυπο ταξινομείται στην Κατηγορία Ω_1

$G(x_{1,} 1, 1) G(x_2, 2, 1) > 2G(x_{1,} 1, 1) G(x_2, 1, 1)$ και $2G(x_{1,} 1, 1) G(x_2, 2, 1) > G(x_{1,} 2, 1) G(x_2, 1, 1)$

το πρότυπο ταξινομείται στην Κατηγορία Ω_2

σε όλες τις άλλες περιπτώσεις το πρότυπο αποθηκεύεται στην Κατηγορία Ω_3

Το ερώτημα (β) λύνεται εύκολα αντικαθιστώντας τις αριθμητικές τιμές του άγνωστου πρότυπου στις σχέσεις του κριτηρίου:

$2G(4,1,1)G(4.5,1,1) > G(4,1,1)G(4.5,2,1)$ και $4G(4,1,1)G(4.5,1,1) > G(4,2,1)G(4.5,1,1)$ το πρότυπο ταξινομείται στην Κατηγορία Ω_1

2*9.39858321892e-11 > 3.79165908855e-08 ψευδές, το πρότυπο **δεν** ταξινομείται στην Κατηγορία Ω_1

$G(4,1,1)G(4.5,2,1) > 2G(4,1,1)G(4.5,1,1)$ και $2G(4,1,1)G(4.5,2,1) > G(4,2,1)G(4.5,1,1)$ το πρότυπο ταξινομείται στην Κατηγορία Ω_2

3.79165908855e-08 > 2*9.39858321892e-11 αληθές και 2*3.79165908855e-08>1.39487342661e-08 αληθές, **το πρότυπο ταξινομείται στην Κατηγορία Ω_2**

Θέμα 3 (4 μονάδες): Εστω νευρωνικό δίκτυο, δύο επιπέδων (ένα κρυφό και το επίπεδο εξόδου) το οποίο αποτελείται από δύο νευρώνες στο κρυφό επίπεδο και έναν νευρώνα στο επίπεδο εξόδου. Το διάνυσμα της εισόδου είναι δισδιάστατο και η συνάρτηση μεταφοράς κάθε νευρώνα είναι της μορφής

$$g(\mathbf{x},\mathbf{w})=\exp(-|\mathbf{w}-\mathbf{x}|^2),$$

όπου **x** είναι το διάνυσμα της εισόδου στον νευρώνα και **w** η "μνήμη" του.
Το πρόβλημα που καλούστε να λύσετε είναι να κατασκευάσετε έναν αλγόριθμο ο οποίος θα εκτιμά τους συντελεστές βαρύτητας των συνάψεων του δικτύου ο οποίος θα βασίζεται στον αλγόριθμο οπισθοδρομικής διάδοσης του σφάλματος.
Η εκπαίδευση θα πρέπει να γίνεται με τέτοιο τρόπο έτσι ώστε το νευρωνικό δίκτυο να αναγνωρίζει τα πρότυπα της μιάς κατηγορίας αν η τιμή της εξόδου είναι μεγαλύτερη του 0.5, σε όλες τις άλλες περιπτώσεις θα αναγνωρίζει πρότυπα της δεύτερης κατηγορίας.

Λύση θέματος 3:
Προφανως πρέπει να κατασκευάσουμε έναν αλγόριθμο που να βασίζεται στην μεθοδολογία της οπισθοδρομικής διάδοσης του σφάλματος.

Γιαυτό τον λόγο θα πρέπει να εξετάσουμε
1. αν η συνάρτηση μεταφοράς του νευρωνικού δικτύου είναι συνεχής και παραγωγίσιμη συνάστηση ως προς τους συντελεστές βαρύτητας των συνάψεων που την απαρτίζουν. Αυτό αποδεικύεται πολύ εύκολα διότι η συνάρτηση μεταφοράς κάθε νευρώνα είναι συνεχής και παραγωγίσιμη και ως προς τις εισόδους και ως προς τους συντελεστές βαρύτητας των συνάψεων. Πρέπει να έχει πεπερασμένα απειροστικά όρια και φραγμένο πεδίο τιμών. Ολες αυτές οι ιδιότητες είναι εύκολο να αποδειχθούν λόγω της υπαρξης της συνάρτησης exp(.)
2. Για να εφαρμόσουμε τον αλγόριθμο οπισθοδρομικής διάδοσης του σφάλματος πρέπει να μοντελοποιήσουμε κάθε νευρώνα σαν μια ακολουθία δυο σειριακά συνδεδεμένων μετασχηματισμών. Ο πρώτος συμπλέκει την πληροφορία της εισόδου με την "μνήμη" του νευρώνα (τους συντελεστές βαρύτητας των συνάψεων) σε μια αριθμητική τιμή η οποία ονομάζεται εσωτερική κατάσταση του νευρώνα, και ο δεύτερος μετασχηματισμός είναι συνήθως μια μη-γραμμική συνάρτηση της εσωτερικής κατάστασης του νευρώνα η οποία δίνει και την αντίστοιχη έξοδο. Μιά απλή λύση στην συνάρτηση μεταφοράς του νευρώνα που δίνεται στην άσκηση είναι:

$$y=|\boldsymbol{w}-\boldsymbol{x}|^2 \quad , \qquad o=\exp(-y)$$

Για τον νευρώνα εξόδου η σχέση εκτίμησης είναι για μεν την ποσότητα δ_ε (σχέση 4.37, σελ. 124 σημειώσεων):

$$\delta_\varepsilon=-\frac{\partial\,\Sigma\varphi\acute{\alpha}\lambda\mu\alpha}{\partial\,o_\varepsilon}\frac{\partial\,o_\varepsilon}{\partial\,y_\varepsilon}=-\frac{\partial(o_\varepsilon-b)^2}{\partial\,o_\varepsilon}\frac{\partial\exp(-y_\varepsilon)}{\partial\,y_\varepsilon}=2(b-o_\varepsilon)\exp(-y_\varepsilon)=2(b-o_\varepsilon)o_\varepsilon$$

όπου o_ε και y_ε είναι αντίστοιχα η έξοδος του νευρώνα εξόδου του δικτύου και η εσωτερική του κατάσταση.

Για τους δύο νευρώνες που βρίσκονται στο επίπεδο εισόδου η αντίστοιχες συναρτήσεις δ_h δίνονται από την σχέση που ακολουθεί (4.41, σελ. 125 σημειώσεων). Η διαγραφή του αθροίσματος πραγματοποιείται γιατί στο επόμενο επίπεδο, το επίπεδο εξόδου, υπάρχει μόνο ένας νευρώνας.

$$\delta_h=-\frac{\partial\,\Sigma\varphi\acute{\alpha}\lambda\mu\alpha}{\partial\,o_h}\frac{\partial\,o_h}{\partial\,y_h}=\frac{\partial\,o_h}{\partial\,y_h}\left(-\frac{\partial\,\Sigma\varphi\acute{\alpha}\lambda\mu\alpha}{\partial\,y_\varepsilon}\right)\frac{\partial\,y_\varepsilon}{\partial\,o_h}=-\exp(-y_h)\,\delta_\varepsilon\frac{\partial(o_h-w_{\varepsilon h})^2}{\partial o_h}=-2\,o_h\delta_\varepsilon(o_h-w_{\varepsilon h})$$

όπου o_h και y_h είναι αντίστοιχα η έξοδος ενός νευρώνα εξόδου του δικτύου και η εσωτερική του κατάσταση και $w_{\varepsilon h}$ είναι αντίστοιχα ο συντελεστής βαρύτητας του νευρώνα εξόδου που φιλτράρει την έξοδο του κρυφού νευρώνα.

Οι συντελεστές βαρύτητας του νευρώνα της εξόδου επαναπροσδιορίζονται από την σχέση:

$$\Delta\,w_{\varepsilon h}=-\alpha\frac{\partial\,\Sigma\varphi\acute{\alpha}\lambda\mu\alpha}{\partial w_{\varepsilon h}}=-\alpha\frac{\partial\,\Sigma\varphi\acute{\alpha}\lambda\mu\alpha}{\partial\,y_\varepsilon}\frac{\partial\,y_\varepsilon}{\partial\,w_{\varepsilon h}}=-a\,\delta_\varepsilon 2(w_{\varepsilon h}-o_h)$$

Οι συντελεστές βαρύτητας των νευρώνων του κρυφού επαναπροσδιορίζονται από την σχέση:

$$\Delta\,w_{hn}=-\alpha\frac{\partial\,\Sigma\varphi\acute{\alpha}\lambda\mu\alpha}{\partial\,w_{hn}}=-\alpha\frac{\partial\,\Sigma\varphi\acute{\alpha}\lambda\mu\alpha}{\partial\,y_h}\frac{\partial\,y_h}{\partial\,w_{hn}}=-a\,\delta_h 2(w_{hn}-x_n)$$

όπου w_{hn} είναι ο συντελεστής βαρύτητας του κρυφού νευρώνα που φιλτράρει την συνιστώσα x_n του διανύσματος εισόδου.

Οι επιθυμητές τιμές για την πρώτη κατηγορία θα είναι b=0 και για την δεύτερη b=1. Κατά την διαδικασία ενεργοποίησης του δικτύου η έξοδος συγκρίνεται με το κατώφλι απόφασης το 0.5. Αν η έξοδος για το διάνυσμα του οποίου δεν γνωρίζουμε την κατηγορία και τίθεται στην είσοδο του νευρωνικού δικτύου παράγει έξοδο μεγαλύτερη του 0.5 τότε το διάνυσμα ταξινομείται στην δεύτερη κατηγορία, αν είναι μικρότερο του 0.5 ταξινομείται στην πρώτη.