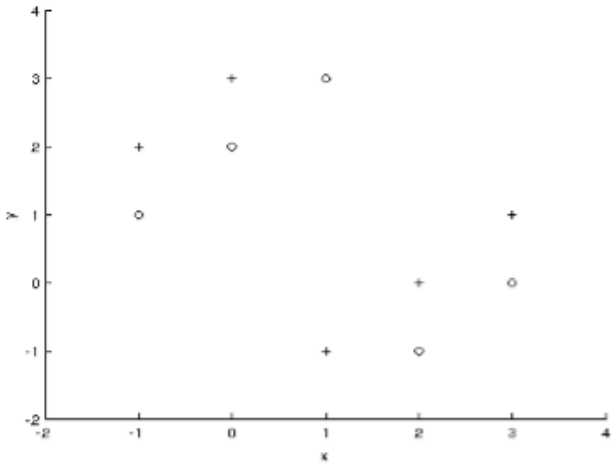
Μηχανική Μάθηση: Ασκήσεις Μελέτης

Τμήμα Πληροφορικής και Επικοινωνιών
Πανεπιστήμιο Πελοποννήσου

Άσκηση 1

Υποθέστε ότι χρησιμοποιούμε τον αλγόριθμο k-κοντινότερων γειτόνων (k-NN) με μέτρο απόστασης την Ευκλείδεια απόσταση στο παρακάτω σύνολο εκπαίδευσης (κάθε σημείο ανήκει είτε στην κλάση (+) ή στην (ο)):

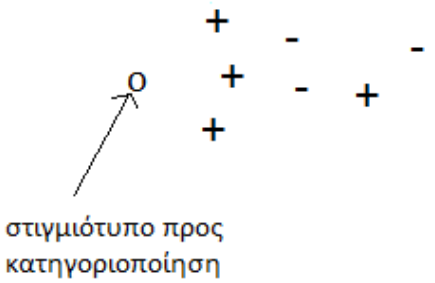


α) Ποιό είναι το 10-fold cross validation σφάλμα αν χρησιμοποιήσουμε τον 1-NN αλγόριθμο για κατηγοριοποίηση;

β) Ποιά από τις παρακάτω τιμές του k οδηγεί στο μικρότερο 10-fold cross validation σφάλμα: 3, 5 ή 9;

Άσκηση 2

Θεωρήστε ότι έχουμε το ακόλουθο σύνολο εκπαίδευσης με θετικά (+) και αρνητικά (-) στιγμιότυπα και ένα νέο στιγμιότυπο (ο) άγνωστης κλάσης.



Αν χρησιμοποιήσουμε τον αλγόριθμο k-κοντινότερων γειτόνων (k-NN) με μέτρο απόστασης την Ευκλείδεια απόσταση, ποιά θα πρέπει να είναι η τιμή του k ώστε το νέο στιγμιότυπο να κατηγοριοποιηθεί στην κλάση (-); Εξηγήστε.

Άσκηση 3

Έστω ότι έχουμε τον παρακάτω πίνακα συνδιακύμανσης:

$$\hat{\mathbf{\Sigma}} = \begin{bmatrix} 1 & 0.5 & 0 \\ 0.5 & 0.333 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Από τον πίνακα αυτό προκύπτουν οι ακόλουθες ιδιοτιμές: $\lambda_1 = 1.267$, $\lambda_2 = 0.065$, $\lambda_3 = 0$ και τα αντίστοιχα ιδιοδιανύσματα:

$$\mathbf{u}_1 = \begin{bmatrix} -0.8817 \\ -0.4718 \\ 0 \end{bmatrix} \qquad \mathbf{u}_2 = \begin{bmatrix} 0.4718 \\ -0.8817 \\ 1 \end{bmatrix} \qquad \mathbf{u}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Έστω ότι εφαρμόζουμε μείωση διάστασης με χρήση του αλγορίθμου PCA. Υπολογίστε τις αναπαραστάσεις των στιγμιοτύπων $\mathbf{x}_1 = (2, 3, 0)^\top$ και $\mathbf{x}_2 = (4, 1, 0)^\top$ στο δισδιάστατο χώρο που ορίζουν οι δυο κύριες συνιστώσες (υποθέστε ότι ο μέσος $\hat{\boldsymbol{\mu}}$ είναι το μηδενικό διάνυσμα, δηλαδή $\hat{\boldsymbol{\mu}} = \mathbf{0}$). Με την προβολή των δεδομένων σε 2 διαστάσεις, πόσο ποσοστό της αρχικής πληροφορίας χάνεται;

Άσκηση 4

Υποθέστε ότι σας δίνεται ένα πρόβλημα δυαδικής κατηγοριοποίησης όπου πρέπει να προβλέψετε την κατηγορία 0 ή 1 χρησιμοποιώντας δυο χαρακτηριστικά x_1, x_2 που παίρνουν πραγματικές τιμές. Μετά από κάποια διαδικασία εκπαίδευσης, μαθαίνετε τον ακόλουθο κανόνα απόφασης:

$$\begin{cases} 1 & \text{αν } w_1 x_1 + w_2 x_2 + w_0 \geq 0 \\ 0 & \text{διαφορετικά} \end{cases}$$

όπου $w_1 = 3$, $w_2 = 5$ και $w_0 = -15$.

(α) Σχεδιάστε το όριο απόφασης και επισημάνετε τις περιοχές όπου θα προβλεφθεί η κλάση 1 και αυτές που θα προβλεφθεί η κλάση 0.

(β) Υποθέστε ότι έχετε μάθει τα παραπάνω βάρη με τη μέθοδο της λογιστικής παλινδρόμησης. Χρησιμοποιώντας αυτό το μοντέλο, ποιά θα ήταν η πρόβλεψή μας για την πιθανότητα ένα στιγμιότυπο να ανήκει στην κλάση 1;

Άσκηση 5

Ο παρακάτω πίνακας δείχνει το αν θα χρησιμοποιήσει ένα άτομο το αυτοκίνητο ανάλογα με την ηλιοφάνεια και τον άνεμο:

Ηλιοφάνεια	Άνεμος	Μετακίνηση
υψηλή	δυνατός	πόδια
υψηλή	μέτριος	πόδια
μέτρια	αδύναμος	πόδια
χαμηλή	δυνατός	αυτοκίνητο
υψηλή	αδύναμος	πόδια
χαμηλή	μέτριος	αυτοκίνητο
μέτρια	δυνατός	αυτοκίνητο

(α) Υπολογίστε την Εντροπία του συνόλου εκπαίδευσης ως προς τη μεταβλητή «Μετακίνηση».

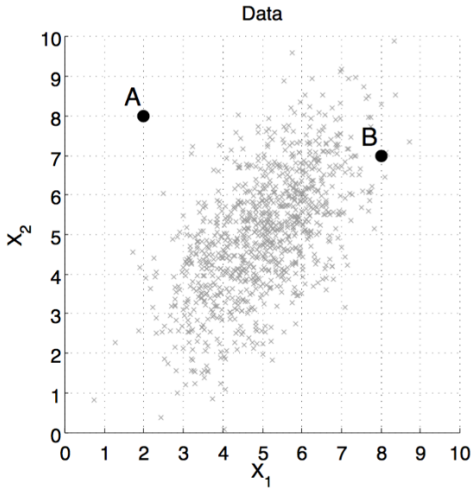
(β) Ποιά μεταβλητή θα χρησιμοποιηθεί σαν ρίζα από τον αλγόριθμο κατασκευής δέντρου απόφασης;

Άσκηση 6

Θεωρείστε ένα πρόβλημα επιβλεπόμενης μάθησης με δεδομένα στον χώρο των 1000 διαστάσεων (δηλαδή 1000 χαρακτηριστικά). Υποθέστε επιπλέον ότι γνωρίζετε ότι μόνο 3 από αυτά τα χαρακτηριστικά είναι σχετικά και βοηθούν στη μάθηση. Ποιόν αλγόριθμο θα χρησιμοποιούσατε; Τον αλγόριθμο k-κοντινότερων γειτόνων (k-NN) ή ένα δέντρο απόφασης; Αιτιολογήσετε την απάντησή σας.

Άσκηση 7

Υποθέστε ότι εφαρμόζουμε τη μέθοδο PCA στα παρακάτω δεδομένα.



Σχεδιάστε τις δυο κύριες συνιστώσες σημειώνοντας ποιά είναι πρώτη και ποιά δεύτερη και περιγράψτε πως καταλήξατε στο συγκεκριμένο αποτέλεσμα. Αν προβάλλουμε τα δεδομένα στην κύρια συνιστώσα, ποιό από τα σημεία A, B θα έχει υψηλότερο σφάλμα ανακατασκευής και γιατι;

Άσκηση 8

Εστω ένα σύνολο δεδομένων παλινδρόμησης $\mathcal{D} = \{(x^{(1)}, y^{(1)}), \ldots, (x^{(N)}, y^{(N)})\}$. Έστω το ακόλουθο μοντέλο:

$$f(x) = w_0$$

(α) Θα θέλαμε να υπολογίσουμε την παραμέτρο w_0 ελαχιστοποιώντας την ακόλουθη συνάρτηση (κόστους) ελαχίστων τετραγώνων:

$$\mathcal{J}(w_0) = \frac{1}{2}\sum_{i=1}^{N}\left(f(x^{(i)}) - y^{(i)}\right)^2$$

Δείξτε ότι το βέλτιστο w_0 είναι:

$$w_0^* = \frac{1}{N}\sum_{i=1}^{N} y^{(i)}$$

(β) Έστω ότι η συνάρτηση κόστους αλλάζει σε:

$$\mathcal{J}_{reg}(w_0) = \frac{1}{2}\sum_{i=1}^{N}\left(f(x^{(i)}) - y^{(i)}\right)^2 + \frac{\lambda}{2}w_0^2$$

όπου $\lambda > 0$. Δείξτε ότι το βέλτιστο w_0 είναι:

$$w_0^* = \frac{1}{N+\lambda}\sum_{i=1}^{N} y^{(i)}$$

Σχολιάστε τι θα συμβεί όταν το λ γίνει πολύ μεγάλο.

Άσκηση 9

Θεωρώντας ότι η πόλωση b έχει ενσωματωθεί στο διάνυσμα των βαρών $\mathbf{w}$ και μια στήλη με τιμή 1 έχει προστεθεί στον πίνακα δεδομένων $\mathbf{X}$, η συνάρτηση κόστους της γραμμικής παλινδρόμησης ορίζεται ως:

$$\mathcal{J}(\mathbf{w}) = \frac{1}{2}\|\mathbf{X}\mathbf{w} - \mathbf{y}\|^2$$

(α) Η γραμμική παλινδρόμηση έχει άμεση λύση η οποία προκύπτει υπολογίζοντας τις μερικές παραγώγους και θέτοντάς τες ίσες με μηδέν. Η άμεση λύση είναι η ακόλουθη:

$$\mathbf{w}^* = (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{y}$$

Δείξτε πώς προκύπτει η παραπάνω λύση. Παρουσιάστε αναλυτικά όλα τα βήματα.

(β) Προσδιορίστε τους κανόνες ενημέρωσης του αλγορίθμου κατάβασης κλίσης για την συνάρτηση κόστους της γραμμικής παλνιδρόμησης. Η απάντησή σας θα πρέπει να δείχνει πως αλλάζει κάθε στοιχείο του διανύσματος $\mathbf{w}$ σε κάθε επανάληψη του αλγορίθμου κατάβασης κλίσης, καθώς και πώς ακριβώς προέκυψε ο κανόνας.

Άσκηση 10

Θεωρήστε ότι εκπαιδεύετε τον αλγόριθμο γραμμικής παλινδρόμησης για είσοδο δυο χαρακτηριστικών και χρησιμοποιείτε ως σύνολο εκπαίδευσης τα παρακάτω δεδομένα:

$$\begin{aligned}
\mathbf{x}_1 &= [1,1]^\top \qquad y_1 = 1 \\
\mathbf{x}_2 &= [3,2]^\top \qquad y_2 = 2 \\
\mathbf{x}_3 &= [4,3]^\top \qquad y_3 = 3 \\
\mathbf{x}_4 &= [5,4]^\top \qquad y_4 = 3 \\
\mathbf{x}_5 &= [6,5]^\top \qquad y_5 = 4
\end{aligned}$$

Υποθέστε ότι ο αλγόριθμος έχει μάθει τα παρακάτω βάρη και πόλωση (παραμέτρους):

$$w_1 = 1 \qquad w_2 = -0.1 \qquad b = 0.1$$

Υπολογίστε το σφάλμα του αλγορίθμου, δηλαδή το μέσο τετραγωνικό σφάλμα.

Άσκηση 11

Έστω ένα μοντέλο το οποίο δεδομένου ενός διανύσματος εισόδου $\mathbf{x}$ το οποίο έχει ετικέτα y, υπολογίζει την απώλεια ως εξής (προς τα εμπρός πέρασμα):

$$\begin{aligned}
z &= \mathbf{w}^\top \mathbf{x} + b \\
\hat{y} &= \sigma(z) \\
\mathcal{L} &= -\big(y \log(\hat{y}) + (1-y)\log(1-\hat{y})\big)
\end{aligned}$$

όπου σ είναι η συνάρτηση sigmoid. Υπολογίστε το προς τα πίσω πέρασμα του δικτύου. Δηλαδή πρέπει να υπολογίσετε την μερική παράγωγο της απώλειας $\mathcal{L}$ ως προς τις παραμέτρους του μοντέλου, δηλαδή τις παρακάτω ποσότητες:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{w}} \qquad \frac{\partial \mathcal{L}}{\partial b}$$

Σημειώστε επίσης ότι η παράγωγος της συνάρτησης $\sigma(x)$ είναι ίση με:

$$\sigma'(x) = \sigma(x)\big(1 - \sigma(x)\big)$$

Άσκηση 12

Έστω ένα αναδρομικό νευρωνικό δίκτυο το οποίο δέχεται ως είσοδο μια ακολουθία από μη αρνητικές τιμες $x^{(1)}, x^{(2)}, \ldots, x^{(T)}$ και υπολογίζει την έξοδο ως εξής:

$$h^{(t)} = \mathrm{ReLU}\Big(w_h\, x^{(t)} + u_h\, h^{(t-1)} + b_h\Big)$$
$$y^{(t)} = \mathrm{ReLU}\Big(w_y\, h^{(t)} + b_y\Big)$$

όπου $w_h, w_y, u_h, b_h, b_y \in \mathbb{R}$ είναι παράμετροι του μοντέλου.

(α) Ορίστε τις τιμές που θα πρέπει να πάρουν οι παράμετροι $w_h, w_y, u_h, b_h, b_y \in \mathbb{R}$ και την αρχική κατάσταση $h^{(0)}$ ώστε το μοντέλο να υπολογίζει στην έξοδο $y^{(T)}$ το άθροισμα των τιμών που δόθηκαν ως είσοδος.

(β) Αν η ακολουθία που δίνεται ως είσοδος στο νευρωνικό δίκτυο είναι η $x^{(1)} = 1$, $x^{(2)} = 8$, $x^{(3)} = 5$, $x^{(4)} = 1$, $x^{(5)} = 2$, $x^{(6)} = 6$, να πραγματοποιηθεί το προς τα εμπρός πέρασμα και να υπολογιστεί η έξοδος $y^{(6)}$. Να δείξετε όλα τα βήματα.

Άσκηση 13

Ο παρακάτω πίνακας δείχνει το αν θα γίνει ένα άτομο αποδεκτό σε ένα συγκεκριμένο Πανεπιστήμιο ανάλογα με το βαθμό που έγραψε στα Μαθηματικά και την Έκθεση:

Μαθηματικά	Έκθεση	Αποδοχή
18	16	Ναι
15	17	Ναι
12	10	Οχι
16	14	Ναι
12	15	Οχι
10	16	Οχι
16	10	Οχι
14	18	Ναι

Έστω ότι εφαρμόζετε τον αλγόριθμο AdaBoost στο παραπάνω σύνολο δεδομένων, κι έστω ότι ο πρώτος αδύναμος αλγόριθμος ταξινόμησης του AdaBoost προβλέπει ότι ένα άτομο γίνεται αποδεκτό αν ο βαθμός του στην Έκθεση είναι μεγαλύτερος από 13.5, διαφορετικά δεν γίνεται αποδεκτό. Να υπολογίσετε το σφάλμα του αλγορίθμου ταξινόμησης err_1, το βάρος του α_1, καθώς και τα νέα βάρη $\mathbf{w}$ των δεδομένων εκπαίδευσης.

Άσκηση 14

Σας δίνονται τα ακόλουθα δεδομένα στον δισδιάστατο χώρο:

Σημείο	x	y
A	1	1
B	1	2
C	2	1
D	1	0
E	8	8
F	10	9
G	10	10

(α) Χρησιμοποιήσετε τον αλγόριθμο k-means με Ευκλείδεια απόσταση για να ομαδοποιήσετε τα δεδομένα. Ας υποθέσουμε ότι τα αρχικά κεντροειδή (centroids) είναι τα σημεία A και F. Εκτελέστε τον αλγόριθμο k-means για δύο επαναλήψεις. Στο τέλος κάθε επανάληψης δείξτε:

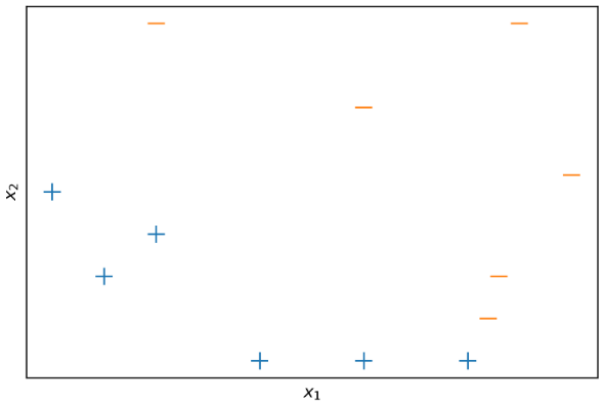
(i) Τις νέες ομάδες (clusters).

(ii) Τα νέα κεντροειδή.

(β) Συγκλίνει ο αλγόριθμος μετά τη δεύτερη επανάληψη; Εάν τα αρχικά κεντροειδή ήταν τα σημεία E και F, πιστεύετε ότι θα χρειαζόταν τον ίδιο αριθμό επαναλήψεων για να συγκλίνει; Εξηγείστε.

Άσκηση 15

Σας δίνονται τα παρακάτω δεδομένα στο χώρο των δυο διαστάσεων (κάθε σημείο ανήκει είτε στην κλάση (+) ή στην (-)):



Έστω ότι εφαρμόζετε τον αλγόριθμο AdaBoost στο παραπάνω σύνολο δεδομένων.

(α) Σχεδιάστε το όριο απόφασης για το δέντρο απόφασης με βάθος ένα το οποίο ελαχιστοποιεί το σταθμισμένο σφάλμα ταξινόμησης. Υποθέστε ότι τα όλα τα στιγμιότυπα έχουν αρχικά ίδια βάρη.

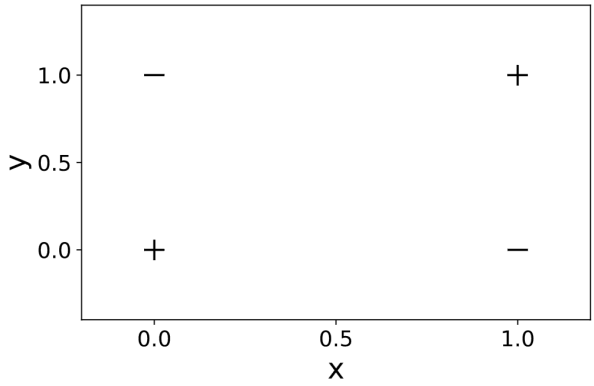
(β) Επιλέξτε ένα αυθαίρετο σημείο που το παραπάνω δέντρο απόφασης ταξινομεί λανθασμένα. Σχεδιάστε έναν κύκλο γύρω από το σημείο που επιλέξατε. Ποιό θα είναι το βάρος του συγκεκριμένου σημείου στην επόμενη επανάληψη του αλγορίθμου AdaBoost;

Άσκηση 16

Σχεδιάζετε ένα σύστημα βαθιάς μάθησης για τη διάγνωση του καρκίνου του θώρακα μέσω ακτινογραφιών. Ποια πιστεύετε ότι θα μπορούσε να είναι η καταλληλότερη μετρική αξιολόγησης: ακρίβεια (accuracy), ευαισθησία (sensitivity) ή ειδικότητα (specificity); Εξηγήστε γιατί.

Άσκηση 17

Σας δίνονται τα παρακάτω δεδομένα στο χώρο των δυο διαστάσεων (κάθε σημείο ανήκει είτε στην κλάση (+) ή στην (-)):



Σχεδιάστε ένα δέντρο απόφασης που ταξινομεί σωστά και τα 4 παραδείγματα του συνόλου δεδομένων.

Άσκηση 18

Σας δίνονται τα παρακάτω δεδομένα εκπαίδευσης για ένα πρόβλημα δυαδικής ταξινόμησης:

x_1	x_2	x_3	y
0	0	0	1
1	0	0	0
0	1	1	0
1	0	1	1

Βρείτε έναν μοντέλο λογιστικής παλινδρόμησης με βάρη w_1, w_2, w_3 και πόλωση b που να ταξινομεί σωστά όλα τα παραδείγματα εκπαίδευσης. Κανένα από τα παραδείγματα δεν πρέπει να βρίσκεται ακριβώς πάνω στο όριο απόφασης.

Άσκηση 19

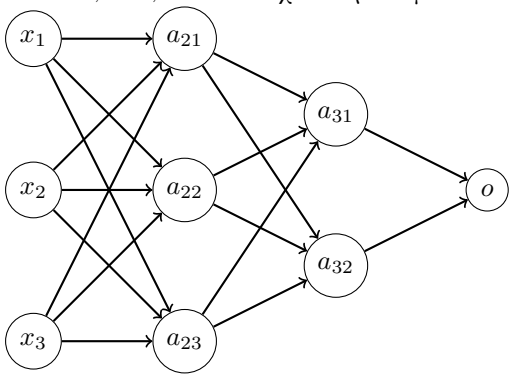
Ένα συνελικτικό επίπεδο σε ένα συνελικτικό νευρωνικό δίκτυο αποτελείται από 8 φίλτρα διάστασης 2×2 (με πολώσεις) και εφαρμόζεται σε μια εικόνα διάστασης $10 \times 10 \times 3$ χωρίς συμπλήρωση μηδενικών (zero padding).

(α) Ποιά είναι η διάσταση της αναπαράστασης της εικόνας στην έξοδο του επιπέδου;

(β) Ποιός είναι ο αριθμός των παραμέτρων (βαρών) που ανανεώνει το μοντέλο κατα τη διάρκεια της εκπαίδευσης;

Άσκηση 20

Σας δίνεται το παρακάτω τεχνητό νευρωνικό δίκτυο μαζί με τις τιμές των παραμέτρων $\mathbf{W}^{(1)}, \mathbf{W}^{(2)}, \mathbf{W}^{(3)}$ και $\mathbf{b}^{(1)}, \mathbf{b}^{(2)}, \mathbf{b}^{(3)}$ που έχουν προκύψει από εκπαίδευση για ένα συγκεκριμένο πρόβλημα.



$$\mathbf{W}^{(1)} = \begin{bmatrix} 0.1 & 0.2 & 0.1 \\ 0.2 & 0.2 & 0.3 \\ 0.2 & 0.1 & 0.1 \end{bmatrix} \quad \mathbf{b}^{(1)} = \begin{bmatrix} 0.3 & 0.1 & 0.1 \end{bmatrix}$$

$$\mathbf{W}^{(2)} = \begin{bmatrix} 0.1 & 0.1 & 0.2 \\ 0.2 & 0.1 & 0.3 \end{bmatrix} \quad \mathbf{b}^{(2)} = \begin{bmatrix} 0.2 & 0.1 \end{bmatrix}$$

$$\mathbf{W}^{(3)} = \begin{bmatrix} 0.1 & 0.3 \end{bmatrix} \quad \mathbf{b}^{(3)} = \begin{bmatrix} 0.1 \end{bmatrix}$$

Υπολογίστε την τιμή στην έξοδο (κόμβος o) για είσοδο $\mathbf{x} = [1, -1, 1]^\top$ δείχνοντας όλα τα ενδιάμεσα βήματα. Θεωρήστε ότι χρησιμοποιούμε τη συνάρτηση ενεργοποίησης ReLU σε όλα τα επίπεδα εκτός από την έξοδο.

Άσκηση 21

Έστω ένα μοντέλο το οποίο δεδομένου ενός διανύσματος εισόδου $\mathbf{x}$ το οποίο έχει ετικέτα y, υπολογίζει την απώλεια ως εξής (προς τα εμπρός πέρασμα):

$$z = w_1 x + w_2 x^2 + b$$
$$\hat{y} = 1 + e^z$$
$$\mathcal{L} = -\frac{1}{2}\big(\log \hat{y} - \log y\big)^2$$

Υπολογίστε το προς τα πίσω πέρασμα του δικτύου. Δηλαδή πρέπει να υπολογίσετε την μερική παράγωγο της απώλειας $\mathcal{L}$ ως προς τις παραμέτρους του μοντέλου, δηλαδή τις παρακάτω ποσότητες:

$$\frac{\partial \mathcal{L}}{\partial w_1} \qquad \frac{\partial \mathcal{L}}{\partial w_2} \qquad \frac{\partial \mathcal{L}}{\partial b}$$