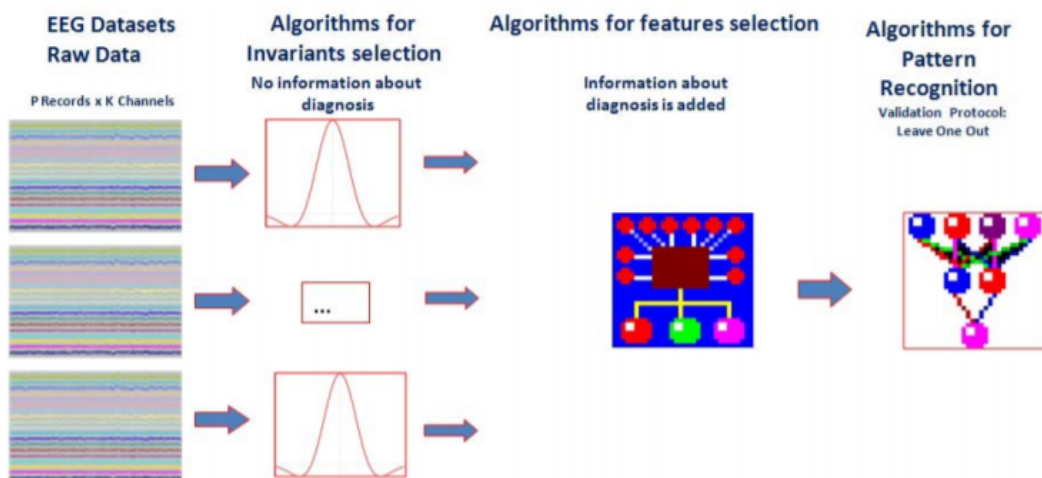


Στατιστική Μοντελοποίηση και Αναγνώριση προτύπων (ΤΗΛ 311)
Αναφορά 3ου Σετ Ασκήσεων

Θέμα 1: Feature Selection – Classification – Cross Validation – Overfitting

Στη παρούσα άσκηση προσομοιώνεται η ανίχνευση του αυτισμού από δεδομένα εγκεφαλογραφήματος(EEG) 25 νεαρών ατόμων, εκ των οποίων οι 15 έχουν διαγνωστεί με αυτισμό ενώ οι 10 δεν έχουν διαγνωστεί με κάποια πάθηση. Δεδομένου των εγκεφαλογραφημάτων υπολογίστηκαν 1000 χαρακτηριστικά, από τα οποία επιλέχθηκαν τα 100 χαρακτηριστικά που διαφοροποιούν καλύτερα τους ασθενείς από τους μη. Λόγω του μικρού αριθμού των δειγμάτων χρησιμοποιήθηκε η τεχνική leave-one-out(cross-validation), κατά την οποία χρησιμοποιείται η ομάδα με τα N-1 δείγματα για να εκπαιδευτεί το μοντέλο και δοκιμάζεται το μοντέλο στο ένα δείγμα της άλλης ομάδας. Η παραπάνω διαδικασία επαναλαμβάνεται N φορές και κάθε φορά αφήνεται εκτός εκπαίδευσης ένα διαφορετικό δείγμα, στο τέλος υπολογίζεται η μέση τιμή των N αποτελεσμάτων.

Σκοπός της προσομοίωσης είναι να διαπιστωθεί σε ποιό σημείο της διαδικασίας είναι ιδανικότερο να εφαρμοστεί η επιλογή των χαρακτηριστικών(feature selection) που διαφοροποιούν καλύτερα τους ασθενείς.



Σχήμα 1 : Διαδικασία ανίχνευσης του αυτισμού

Κατά τη προσομοίωση του πειράματος παράγονται με εντελώς τυχαίο τρόπο 1000 χαρακτηριστικά για 25 άτομα(15 ασθενείς και 10 μη ασθενείς) και πραγματοποιείται ταξινόμηση των δειγμάτων με 3 διαφορετικές τεχνικές :

- Classify without feature selection
- Classify with feature selection inside the cross validation
- Classify with feature selection outside the cross validation

Η υλοποίηση των προαναφερθέντων τεχνικών πραγματοποιείται στο αρχείο `overfittingTest.m` και η ακρίβεια ταξινόμησης των παραπάνω τεχνικών παρουσιάζεται στο παρακάτω πίνακα.

	Without feature Selection	Inside the cross validation	Outside the cross validation
Accuracy	0.6	0.52	1.00

Πίνακας 1.1 : Ακρίβεια ταξινόμησης

Με την απουσία εφαρμογής του feature selection διαπιστώνεται ακρίβεια της τάξης του 60%. Εφαρμόζοντας όμως feature selection παρατηρείται σημαντική μεταβολή στην ακρίβεια του μοντέλου ανάλογα αν το feature selection θα πραγματοποιηθεί μέσα στο cross-validation ή έξω από το cross-validation. Συγκεκριμένα χρησιμοποιώντας τη τεχνική feature selection outside the cross validation παρατηρείται ότι η ακρίβεια του μοντέλου είναι 100% δηλαδή το φαινόμενο overfitting καθώς το feature selection πραγματοποιείται σε 24 δείγματα εκπαίδευσης και σε 1 δείγμα ελέγχου. Αντιθέτως με τη χρήση της τεχνικής feature selection inside the cross validation η ακρίβεια του μοντέλου κυμαίνεται γύρω στο 55% και αντικατοπτρίζει τη πραγματικότητα λόγω της τυχαιότητας των δειγμάτων.

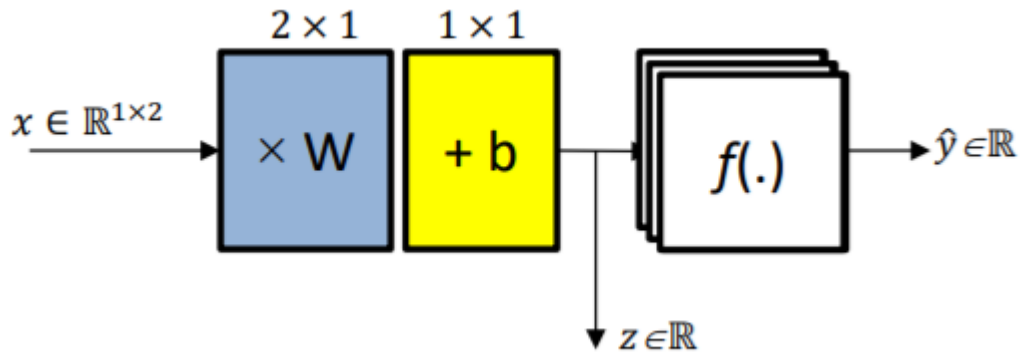
```
Classify without feature selection  
accuracy :0.6
```

```
Classify with feature selection inside the cross validation  
accuracy :0.52
```

```
Classify with feature selection outside the cross validation  
accuracy :1
```

Θέμα 2: Υλοποίηση ενός απλού νευρωνικού δικτύου

Στη 2η άσκηση υλοποιήθηκε και εκπαιδεύτηκε ένα νευρωνικό δίκτυο χωρίς τη χρήση έτοιμων προγραμμάτων και βιβλιοθηκών.



Σχήμα 2 : Block Diagram του νευρωνικού Δικτύου

Ως σύνολο δεδομένων εκπαίδευσης του νευρωνικού δικτύου ορίζεται $D = \{(x^{(1)}, y^{(1)}), \dots, (x^{(N)}, y^{(N)})\}$ με $x^{(i)} \in \mathbb{R}^{1 \times 2}$ και $y^{(i)} \in \mathbb{R}$.

Το νευρωνικό δίκτυο ορίζεται από τη σχέση:

$$\hat{y}^{(i)} = f(x^{(i)}W + b), \text{ όπου } x^{(i)} \in \mathbb{R}^{1 \times 2} \text{ είναι ένα δείγμα, } W \in \mathbb{R}^{2 \times 1}, b \in \mathbb{R}^{1 \times 1} \text{ και } f(z) = \frac{1}{1+e^{-z}}.$$

Το σφάλμα ανάμεσα στη πρόβλεψη $\hat{y}^{(i)}$ του νευρωνικού δικτύου και στην τιμή $y^{(i)}$ που αντιστοιχεί στα δεδομένα εισόδου $x^{(i)}$, υπολογίζεται χρησιμοποιώντας τη συνάρτηση cross entropy

$$J(y^{(i)}, \hat{y}^{(i)}; W, b) = -y^{(i)} \ln(\hat{y}^{(i)}) - (1 - y^{(i)}) \ln(1 - \hat{y}^{(i)})$$

Για να επιτευχθεί αριθμητική ευστάθεια, η συνάρτηση cross-entropy υπολογίζεται ως ο μέσος όρος από ένα σύνολο B δειγμάτων (batch).

$$J(Y, \hat{Y}; W, b) = \frac{1}{B} \sum_i (-y^{(i)} \ln(\hat{y}^{(i)}) - (1 - y^{(i)}) \ln(1 - \hat{y}^{(i)})) \quad (1)$$

α) Θέτοντας $z^{(i)} = x^{(i)}W + b$ οπότε $\hat{y}^{(i)} = f(z^{(i)})$

$$\ln(\hat{y}^{(i)}) = \ln(f(z^{(i)})) = \ln\left(\frac{1}{1+e^{-z^{(i)}}}\right) = -\ln(1 + e^{-z^{(i)}})$$

$$\ln(1 - \hat{y}^{(i)}) = \ln(1 - f(z^{(i)})) = \ln\left(1 - \frac{1}{1+e^{-z^{(i)}}}\right) = \ln\left(\frac{e^{-z^{(i)}}}{1+e^{-z^{(i)}}}\right) = \ln(e^{-z^{(i)}}) - \ln(1 + e^{-z^{(i)}})$$

Από την (1) προκύπτει ότι :

$$(1) \Rightarrow J(Y, \hat{Y}; W, b) = \frac{1}{B} \sum_i (y^{(i)} \ln(1 + e^{-z^{(i)}}) - (1 - y^{(i)}) (\ln(e^{-z^{(i)}}) - \ln(1 + e^{-z^{(i)}}))) \Rightarrow$$

$$J(Y, \hat{Y}; W, b) = \frac{1}{B} \sum_i (-\ln(e^{-z^{(i)}}) + y^{(i)} \ln(e^{-z^{(i)}}) + \ln(1 + e^{-z^{(i)}})) \Rightarrow$$

$$J(Y, \hat{Y}; W, b) = \frac{1}{B} \sum_i (z^{(i)} - y^{(i)} z^{(i)} + \ln(1 + e^{-z^{(i)}})) \quad (2)$$

β)

Παραγωγίζοντας την εξίσωση (2) του προηγούμενου ερωτήματος :

$$\frac{\partial J}{\partial z^{(i)}} = \frac{dz^{(i)}}{dz^{(i)}} - \frac{d(y^{(i)} z^{(i)})}{dz^{(i)}} + \frac{d(\ln(1+e^{-z^{(i)}}))}{dz^{(i)}} = 1 - y^{(i)} - \frac{e^{-z^{(i)}}}{1+e^{-z^{(i)}}} = -y^{(i)} + \frac{1+e^{-z^{(i)}} - e^{-z^{(i)}}}{1+e^{-z^{(i)}}} = -y^{(i)} + \frac{1}{1+e^{-z^{(i)}}}$$

$$\frac{\partial J}{\partial z^{(i)}} = -y^{(i)} + \hat{y}^{(i)} \quad \forall i \in B$$

γ)

Με τον κανόνα της αλυσίδας υπολογίζονται οι μερικοί παράγωγοι :

$$\frac{dJ}{dW} = \frac{dJ}{dz^{(i)}} \cdot \frac{dz^{(i)}}{dW} = (-y^{(i)} + \hat{y}^{(i)}) \cdot (x^{(i)})^T$$

$$\frac{dJ}{db} = \frac{dJ}{dz^{(i)}} \cdot \frac{dz^{(i)}}{db} = -y^{(i)} + \hat{y}^{(i)}$$

δ)

Τέλος συμπληρώνονται τα κατάλληλα αρχεία Matlab προκειμένου να υλοποιηθεί το παραπάνω νευρωνικό δίκτυο. Συγκεκριμένα συμπληρώνονται τα αρχεία “forward.m” και “sigmoid.m” ώστε να υπολογιστεί η ενδιάμεση έξοδος $z = x^{(i)}W + b$ και να πραγματοποιηθεί η εκτίμηση της χρησιμοποιώντας τη σιγμοειδής συνάρτηση. Επιπλέον για το συγκεκριμένο batch υπολογίζεται το κόστος υλοποιώντας τη συνάρτηση “cross_entropy.m”. Τέλος επαναλαμβάνεται η παραπάνω διαδικασία χρησιμοποιώντας τα νέα βάρη και biases τα οποία υπολογίζονται ως εξής:

$$W = W - (\text{learning rate}) \frac{dJ}{dW}$$

$$b = b - (\text{learning rate}) \frac{dJ}{db}$$

Το learning rate ισούται με 0.001 και οι μερικές παράγωγοι ως προς W,b υπολογίζονται από τη συνάρτηση “backward.m”.

Μετά την εκπαίδευση του νευρωνικού δικτύου παρατηρείται ότι η ακρίβεια ταξινόμησης του στα δείγματα ελέγχου είναι γύρω στο 90%.

Θέμα 3: Convolutional Neural Networks for Image Recognition

Στη τρίτη άσκηση υλοποιούνται διαφοροί ταξινομητές νευρωνικών δικτύων για την αναγνώριση φωτογραφιών ρούχων χρησιμοποιώντας ως σύνολο δεδομένων τις εικόνες του Fashion-MNIST οι οποίες χωρίζονται σε 10 κατηγορίες {'T-shirt/top', 'Trouser', 'Pullover', 'Dress', 'Coat', 'Sandal', 'Shirt', 'Sneaker', 'Bag', 'Ankle boot'}.

1)

Αρχικά υλοποιείται ένας ταξινομητής με dense neural networks συμπληρώνοντας τον απαραίτητο κώδικα fashion.py και εκπαιδευεται το νευρωνικό δίκτυο για 400 εποχές με χρήση του αλγορίθμου βελτιστοποίησης adam.

Οι ακρίβεια του training και του validation απεικονίζονται παρακάτω :

Training Accuracy	Validation Accuracy
0.9957	0.8859

2)