



#### Οδηγίες:

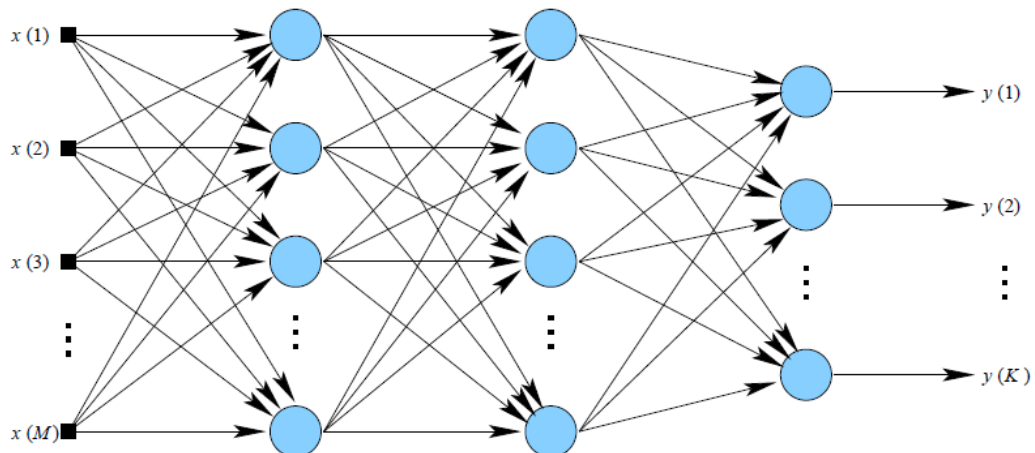
1. Σας παρακαλώ να σεβαστείτε τον παρακάτω κώδικα τιμής τον οποίον θα θεωρηθεί ότι προσυπογράφετε μαζί με τη συμμετοχή σας στο μάθημα και τις εργασίες του:
  - a) Οι απαντήσεις στις εργασίες, τα quiz και τις εξετάσεις, ο κώδικας και γενικά οτιδήποτε αφορά τις εργασίες θα είναι προϊόν δικής μου δουλειάς.
  - b) Δεν θα διαθέσω κώδικα, απαντήσεις και εργασίες μου σε κανέναν άλλο.
  - c) Δεν θα εμπλακώ σε άλλες ενέργειες με τις οποίες ανέντιμα θα βελτιώνω τα αποτελέσματα μου ή ανέντιμα θα αλλάζω τα αποτελέσματα άλλων.
2. Η εργασία είναι ατομική
3. Ημερομηνία παράδοσης: **Τετάρτη, 9/5/2018 στις 20:00**
4. **Παραδοτέα:** α) Κώδικας και β) Αναφορά με τις απαντήσεις, παρατηρήσεις, πειράματα, αποτελέσματα και οδηγίες χρήσης του κώδικα.

#### Θέμα 1: Αρχιτεκτονική Νευρωνικών δικτύων

Εξετάστε την παραδοσιακή αρχιτεκτονική ρηχών νευρωνικών δικτύων. Συγκεκριμένα, έστω το νευρωνικό δίκτυο του παρακάτω σχήματος. Ας υποθέσουμε ότι η είσοδος του είναι RGB εικόνες σημάτων κυκλοφορίας.

Η δομή του δικτύου είναι η εξής:

- Είσοδος:  $3 \times 64 \times 64 = 12288$  – διάστατη
- Στο πρώτο επίπεδο υπάρχουν 100 κόμβοι (εμφανίζονται με γαλάζιο χρώμα στο σχήμα)
- Στο δεύτερο επίπεδο υπάρχουν 100 κόμβοι (γαλάζιο χρώμα)
- Στο τρίτο επίπεδο υπάρχουν 10 κόμβοι (γαλάζιο χρώμα, ένας για κάθε κλάση)



Υπολογίστε τον συνολικό αριθμό των παραμέτρων (βάρη) στο παραπάνω δίκτυο. Αιτιολογήστε την απάντησή σας.

## Θέμα 2: Εκτίμηση Παραμέτρων (Maximum Likelihood)

Υποθέστε ότι  $n$  δείγματα  $D = \{x_1, \dots, x_n\}$  παράγονται ανεξάρτητα από μια κατανομή Poisson με παράμετρο  $\lambda$ :

$$p(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \dots, \quad \lambda > 0$$

Βρείτε τον εκτιμητή μεγίστης πιθανοφάνειας της παραμέτρου  $\lambda$ .

## Θέμα 3: Εκτίμηση Παραμέτρων και Ταξινόμηση (Maximum Likelihood - Naïve Bayes Classifier)

Σε αυτή την άσκηση θα υλοποιήσετε τον naive Bayes ταξινομητή για αναγνώριση ψηφίων. Τα δεδομένα της άσκησης υπάρχουν στο αρχείο `digits.mat`. Υπάρχουν 10 κλάσεις που αντιστοιχούν στα ψηφία από το 0 έως το 9. Κάθε πίνακας χαρακτηριστικών είναι μια ασπρόμαυρη εικόνα με διαστάσεις  $28 \times 28$  ο οποίος μπορεί να αναπαρασταθεί σαν ένα διάνυσμα  $784 \times 1$  με στοιχεία τις τιμές 0 ή 1.

Τα δείγματα εκπαίδευσης είναι στους πίνακες `train0, ..., train9`. Αντίστοιχα τα δείγματα δοκιμής είναι στους πίνακες `test0, ..., test9`. Σε κάθε πίνακα, κάθε γραμμή είναι ένα διάνυσμα το οποίο μπορεί να γίνει μία  $28 \times 28$  ασπρόμαυρη εικόνα για οπτικοποίηση.

```
load('digits');  
A = reshape(train3(43, :), 28, 28)';  
imagesc(A);  
Colorbar
```

Στον naive Bayes ταξινομητή, υποθέτουμε ότι τα χαρακτηριστικά, τα οποία στην περίπτωση μας, είναι εικονοστοιχεία (pixel) των εικόνων, είναι ανεξάρτητα. Θα χρησιμοποιήσουμε την κατανομή Bernoulli για να μοντελοποιήσουμε το κάθε χαρακτηριστικό (pixel) κάθε κλάσης.

Έστω  $x^{y_i}$  ότι είναι η τυχαία μεταβλητή που παριστάνει το  $i$ -στο χαρακτηριστικό του ψηφίου  $y$  και  $p^{y_i}$  η αντίστοιχη παράμετρος της κατανομής Bernoulli του  $i$ -στού χαρακτηριστικού του ψηφίου  $y$ .

1. Βρείτε τον εκτιμητή μεγίστης πιθανοφάνειας της παραμέτρου  $p^{y_i}$  δοσμένων  $n$  δειγμάτων  $\{x_1^{y_i}, x_2^{y_i}, \dots, x_n^{y_i}\}$
2. Χρησιμοποιήστε τους εκτιμητές μεγίστης πιθανοφάνειας για να εκπαιδεύσετε μοντέλα για κάθε ψηφίο χρησιμοποιώντας τα δοσμένα δείγματα εκπαίδευσης. Οπτικοποιήστε τα εκπαιδευμένα μοντέλα για κάθε ψηφίο σχηματίζοντας για κάθε ψηφίο  $28 \times 28$  εικόνες των οποίων η φωτεινότητα κάθε εικονοστοιχείου αναπαριστά την παράμετρο  $p^{y_i}$  της κατανομής Bernoulli που αντιστοιχεί στο  $i$ -στο εικονοστοιχείο του ψηφίου  $y$ .
3. Χρησιμοποιήστε τα εκπαιδευμένα μοντέλα για να ταξινομήσετε τα ψηφία στο σύνολο δοκιμής (test set). Βρείτε την ακρίβεια της ταξινόμησης (classification accuracy) ως τον λόγο των ψηφίων που έχουν ταξινομηθεί σωστά ως προς το σύνολο των δειγμάτων ελέγχου. Υπολογίστε τον πίνακα confusion matrix που το στοιχείο του  $(i, j)$  παριστάνει πόσο συχνά η εικόνα του ψηφίου  $i$  ταξινομείται ως ψηφίο  $j$ .

#### Θέμα 4: Bayesian εκτίμηση παραμέτρου (Bayesian Inference)

Θεωρήστε ένα πείραμα όπου κάνουμε  $n$  μετρήσεις στις χρονικές στιγμές  $k = 1, 2, \dots, n$ . Στο αρχείο data\_2\_4.txt δίνεται ένα σύνολο  $H(n) = \{x_k, k = 1, \dots, n\}$ ,  $n$  μετρήσεων, όπου  $n = 25$ . Υποθέτουμε ότι η συνάρτηση πυκνότητας πιθανότητας των μετρήσεων είναι Γκαουσιανή  $p(x) \sim \mathcal{N}(\mu, \sigma^2)$ , με γνωστή τυπική απόκλιση  $\sigma = 1.25$ . Επίσης, υποθέτουμε ότι έχουμε κάποια εκ των προτέρων πληροφορία για την κατανομή της μέσης τιμής  $\mu$ :  $p(\mu) \sim \mathcal{N}(\mu_0, \sigma_0^2)$ .

Σε αυτή την άσκηση, μελετούμε την Bayesian εκτίμηση του  $\mu$ .

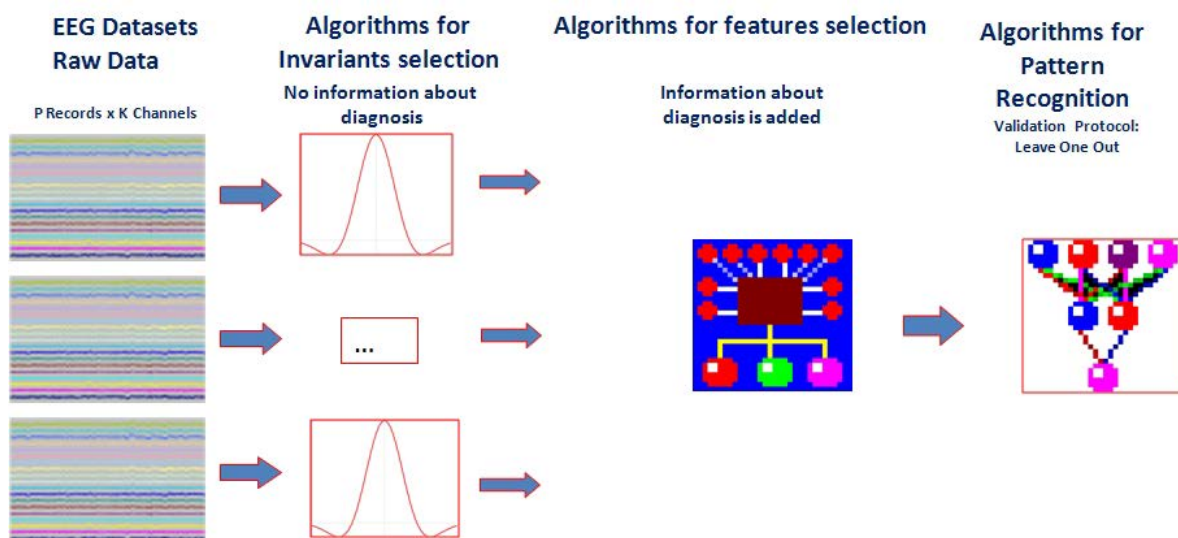
α) Υποθέτοντας ότι  $\mu_0 = 0$  και  $\sigma_0^2 = 10\sigma^2$ , ζωγραφίστε στο ίδιο σχήμα την δεσμευμένη πυκνότητα  $p(\mu|H(n))$ , καθώς το  $n$  μεταβάλλεται από 1 έως 25.

β) Σχεδιάστε τις δεσμευμένες πυκνότητες  $p(x|H(n))$  που προκύπτουν από την Bayesian εκτίμηση όταν  $\sigma^2 = 10\sigma^2$ ,  $\sigma_0^2 = \sigma^2$ ,  $\sigma_0^2 = 0.1\sigma^2$  και  $\sigma_0^2 = 0.01\sigma^2$ . Τι παρατηρείτε;

#### Θέμα 5: Feature Selection – Classification – Cross Validation – Overfitting

Σε επιστημονικό περιοδικό που δημοσιεύει πρωτότυπες εργασίες στον τομέα της βιοϊατρικής πληροφορικής, στάλθηκε ένα άρθρο για αυτόματη ανίχνευση του αυτισμού από δεδομένα εγκεφαλογραφήματος (EEG). Οι ερευνητές χρησιμοποίησαν EEG δεδομένα από ένα σύνολο 25 νεαρών ατόμων, 15 από τους οποίους είχαν διαγνωστεί με αυτισμό και 10 ήταν φυσιολογικοί (χωρίς την διάγνωση κάποιας πάθησης). Από τα εγκεφαλογραφήματα οι ερευνητές υπολόγισαν έναν μεγάλο αριθμό χαρακτηριστικών (περίπου 1000). Λόγω του μεγάλου αριθμού των χαρακτηριστικών οι ερευνητές διάλεξαν τα 100 χαρακτηριστικά που διαφοροποιούν καλύτερα τους ασθενείς από τους φυσιολογικούς. Επίσης λόγω του μικρού αριθμού των δειγμάτων χρησιμοποίησαν την τεχνική leave-one-out (cross validation). Σύμφωνα με αυτή την τεχνική αν έχουμε  $N$  δείγματα, τα χωρίζουμε σε δύο ομάδες  $N-1$  δειγμάτων και ενός δείγματος. Χρησιμοποιούμε την ομάδα με τα  $N-1$  δείγματα για να εκπαιδεύσουμε το μοντέλο και δοκιμάζουμε το μοντέλο στο ένα δείγμα της άλλης ομάδας. Επαναλαμβάνουμε αυτή την διαδικασία  $N$  φορές και κάθε φορά αφήνουμε εκτός εκπαίδευσης ένα διαφορετικό δείγμα. Στο τέλος παίρνουμε την μέση τιμή των  $N$  αποτελεσμάτων.

Οι συγγραφείς του άρθρου οργάνωσαν την έρευνά τους σύμφωνα με το παρακάτω διάγραμμα.



Ένας από τους κριτές διαμαρτυρήθηκε ότι το cross-validation πρέπει να περιλαμβάνει και το feature selection (diagnosis added). Στην άσκηση αυτή θα δούμε αν είχε δίκαιο, συμπληρώνοντας τον κώδικα overfittingTest.m.

Στο κώδικα αυτό παράγουμε με εντελώς τυχαίο τρόπο 1000 χαρακτηριστικά για 25 άτομα (15 ασθενείς και 10 φυσιολογικοί). Όπως είναι αναμενόμενο τα τυχαία χαρακτηριστικά δεν μας δίνουν καμία πληροφορία για ένα άτομο και ένας ταξινομητής που θα εκπαιδευτεί με αυτά τα δεδομένα θα προβλέπει την κατηγορία ενός ατόμου με πιθανότητα κοντά στο 50%.

Τρέξτε το τμήμα του κώδικα «Classify without feature selection» για να επιβεβαιώσετε τα παραπάνω. Αν δεν είναι διαθέσιμα τα Support Vector Machines (svmtrain, svmclassify) στον υπολογιστή που εργάζεστε τότε χρησιμοποιήστε ένα οποιονδήποτε άλλο ταξινομητή (π.χ. Linear and Quadratic Discriminant Analysis)

```
lda = fitcdiscr(data, labels);
ldaClass = resubPredict(lda);

qda = fitcdiscr(data, labels, 'DiscrimType', 'quadratic');
qdaResubErr = resubLoss(qda);
```

Τώρα θα δοκιμάσουμε να κάνουμε feature selection. Μια απλή μέθοδος είναι να διαλέξουμε τα χαρακτηριστικά που σχετίζονται καλύτερα με τα labels.

```
% Feature selection
r = zeros(numFeatures, 1);
for j = 1:numFeatures
    % Compute the similarity measure
    r(j) = similarityMeasure(features, labels);
end
[rSorted, sortedFeatureIndices] = sort(r, 'descend');
% The indices of the selected features.
selectedIndices = sortedFeatureIndices(1:numSelectedFeatures);
```

- A) Συμπληρώσετε τον κώδικα όπου το feature selection γίνεται 25 φορές μέσα στο cross-validation (με τα 24 δείγματα εκπαίδευσης).
- B) Συμπληρώσετε τον κώδικα όπου το feature selection γίνεται μία φορά πριν το cross-validation (με όλα τα 25 δείγματα).