ΗΜΜΥ ΕΜΠ

Αναγνώριση Προτύπων (2022-2023)

Προπαρασκευή 2ης Εργαστηριακής Άσκησης

Ηλιόπουλος Γεώργιος:	03118815	giliopoulos301@gmail.com
Σερλής Εμμανουήλ Αναστάσιος:	03118125	manosserlis@gmail.com

Θέμα: Αναγνώριση φωνής με Κρυφά Μαρκοβιανά Μοντέλα και Αναδρομικά Νευρωνικά Δίκτυα

Σκοπός είναι η υλοποίηση ενός συστήματος επεξεργασίας και αναγνώρισης φωνής, με εφαρμογή σε αναγνώριση μεμονωμένων λέξεων. Στο στάδιο της προπαρασκευής θα γίνει εξαγωγή κατάλληλων ακουστικών χαρακτηριστικών από φωνητικά δεδομένα, χρησιμοποιώντας τα κατάλληλα πακέτα python, καθώς και ανάλυση και απεικόνισή τους με σκοπό την κατανόηση και την εξαγωγή χρήσιμων πληροφοριών από αυτά.

Βήμα 1: Ανάλυση αρχείων ήχου με το Praat

Τα χαρακτηριστικά που εξήχθησαν κάνοντας χρήση του λογισμικού Praat είναι τα ακόλουθα:

Αρχείο	Φωνήεν	Pitch(Hz)
onetwothree1.wav	'α'	135.2
onetwothree1.wav	'ου'	129.9
onetwothree1.wav	'ι'	131.08
onetwothree8.wav	'α'	180.06
onetwothree8.wav	'ου'	190.09
onetwothree8.wav	'ι'	179.58

Πίνακας 1

Αρχείο	Φωνήεν	Formant1(Hz)	Formant2(Hz)	Formant3(Hz)
onetwothree1.wav	'α'	706.52	1102.28	2368
onetwothree1.wav	'ου'	433.85	1834.82	2491.39
onetwothree1.wav	'ι'	394.74	1845.37	3232.95
onetwothree8.wav	'α'	713.17	1407.42	2827.28
onetwothree8.wav	'ου'	343.56	1754.9	2702.77
onetwothree8.wav	'ι'	377.9	2113.15	2750.97

Πίνακας 2

Βήμα 2: Data Parser

Σε αυτό το βήμα φτιάξαμε μία συνάρτηση data parser που διαβάζει το σύνολο των δεδομένων μας και επιστρέφει το wav, τον ομιλητή και το ψηφίο. Αφού τρέξουμε την συνάρτηση διαπιστώνουμε πως έχουμε 133 wav αρχεία ήχου για 9 ψηφία $(1-9)$ από 15 διαφορετικούς εκφωνητές.

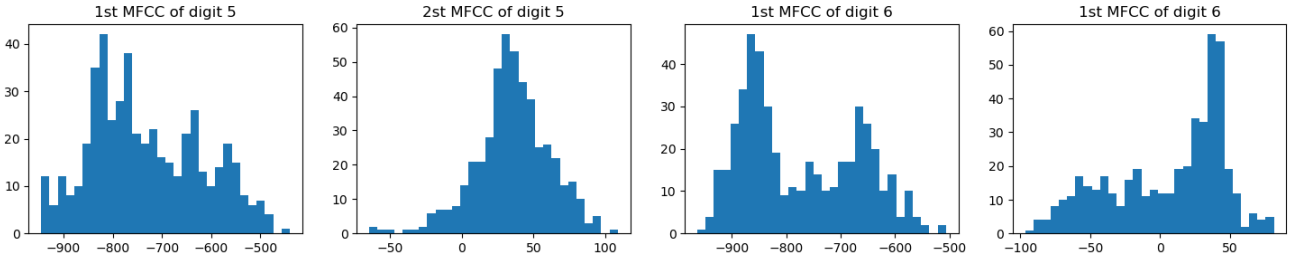
Βήμα 3: Mel-Frequency Cepstral Coefficients (MFCCs)

Για την εξαγωγή των mfccs χρησιμοποιήθηκαν παράθυρα $25ms$ και βήμα $10ms$. Επίσης υπολογίστηκαν η πρώτη και η δεύτερη τοπική παράγωγος των χαρακτηριστικών deltas και delta-deltas. Για κάθε wav αρχείο δημιουργήθηκε ένα διάνυσμα 13×30 που υποδηλώνει τα 13 χαρακτηριστικά για τα 30 παράθυρα. Τα διανύσματα αυτά αποθηκεύτηκαν σε μία λίστα μήκους 133, όσες δηλαδή και οι εκφωνήσεις.

Βήμα 4: Ιστογράμματα MFCCs και Mel Filterbank Spectral Coefficients (MFSCs)

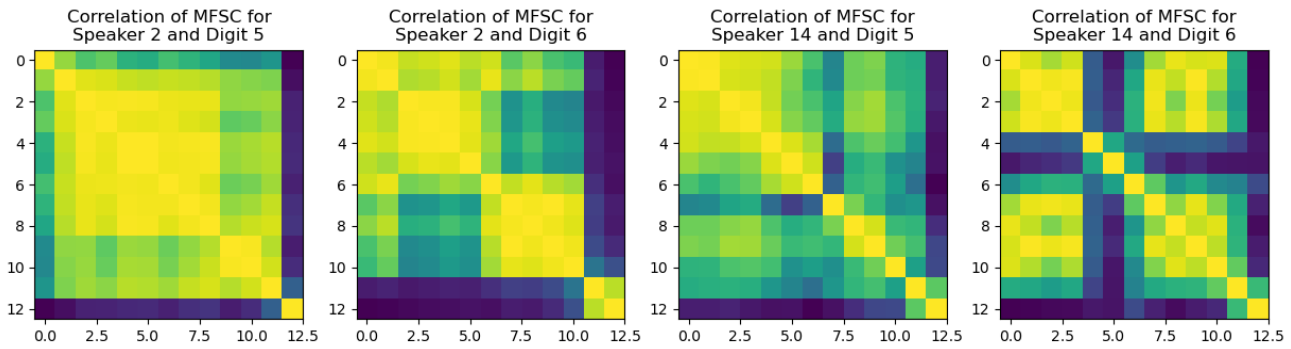
Επειδή ο αριθμός μητρώου και των δύο μας τελειώνει σε 5 πήραμε ως $n1$ και $n2$ τα ψηφία 5 και 6 αντίστοιχα.

Στην εικόνα 1 απεικονίζουμε τα ιστογράμματα των πρώτων 2 mfccs των ψηφίων 5 και 6 από όλες τις εκφωνήσεις. Παρατηρούμε πως αν και ως άκουσα φωνητικά οι λέξεις «five» και «eight» διαφέρουν αρκετά τα ιστογράμματα των πρώτων 2 mfcc χαρακτηριστικών διαφέρουν ελάχιστα. Αντιλαμβανόμαστε λοιπόν πως για ένα μοντέλο φωνητικής αναγνώρισης ψηφίων απαιτούνται και τα υπόλοιπα χαρακτηριστικά που υπολογίσαμε στο βήμα 3.

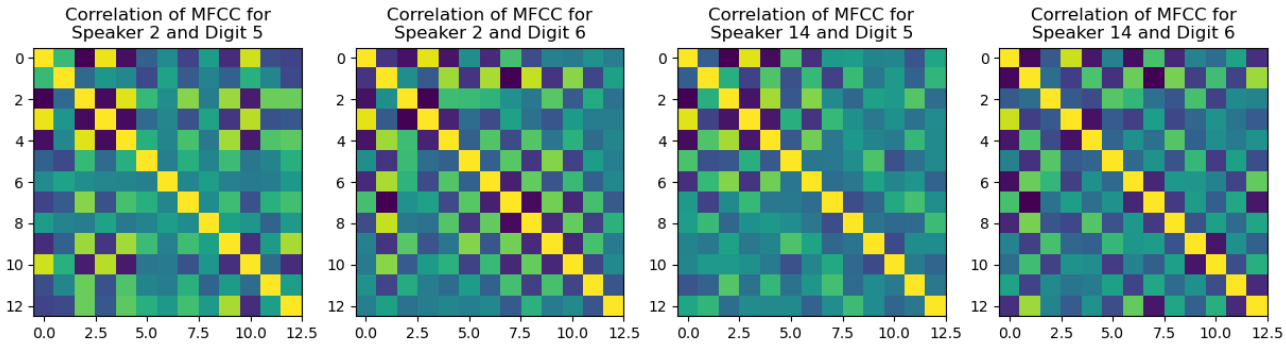


Εικόνα 1: Ιστογράμματα 1ου και 2ου mfcc χαρακτηριστικού των ψηφίων 5 και 6

Στη συνέχεια μελετήσαμε τα Mel Filterbank Spectral Coefficients (MFSCs), δηλαδή τα χαρακτηριστικά που εξάγονται αφού εφαρμοστεί η συστοιχία φίλτρων της κλίμακας Mel πάνω στο φάσμα του σήματος φωνής αλλά χωρίς να εφαρμοστεί στο τέλος ο μετασχηματισμός DCT. Στην εικόνα 2 παρουσιάζουμε την συσχέτιση των mfsc δύο διαφορετικών εκφωνήσεων από το ψηφίο 5 και 6 (4 εκφωνήσεις σύνολο). Για σκοπούς σύγκρισης στην εικόνα 3 βλέπουμε την συσχέτιση των mfcc των ίδιων εκφωνήσεων.



Εικόνα 2: Συσχέτιση MFSC των ψηφίων 5 και 6 από τους εκφωνητές 2 και 14



Εικόνα 3: Συσχέτιση MFCC των ψηφίων 5 και 6 από τους εκφωνητές 2 και 14

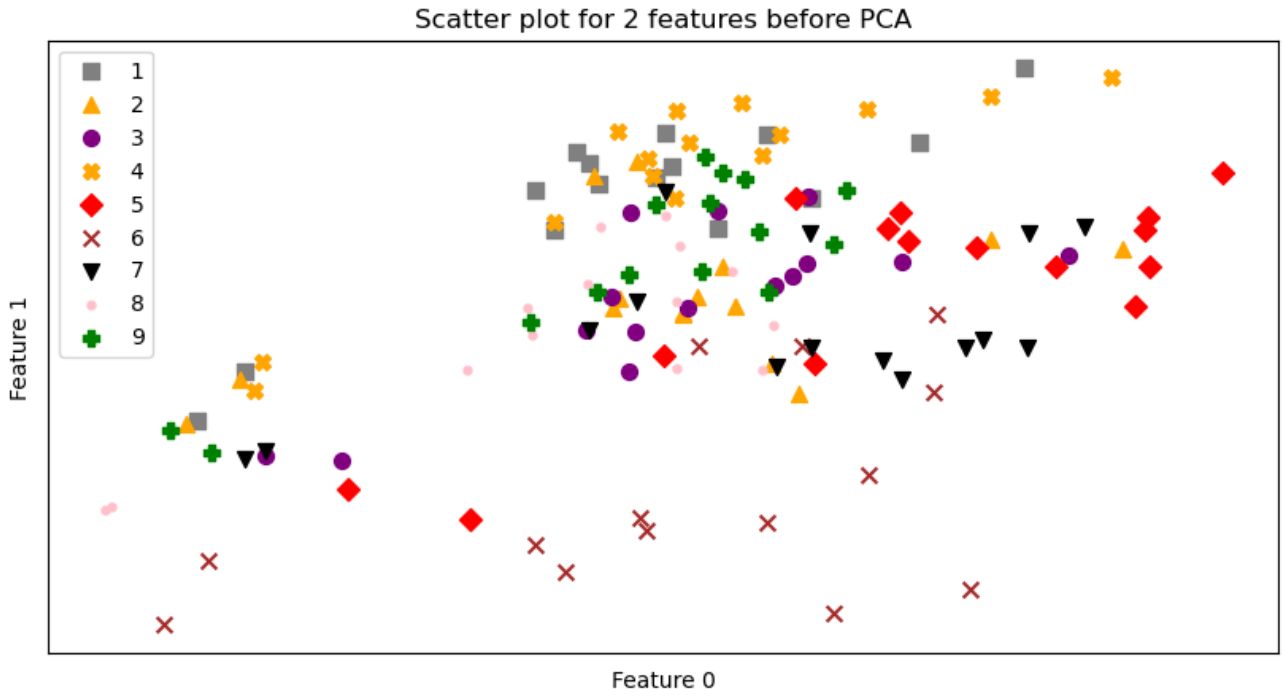
Συγκρίνοντας τις εικόνες 2 και 3 διαπιστώνουμε πως υπάρχει αρκετά υψηλή συσχέτιση στα MFSC χαρακτηριστικά αφού παρατηρούνται υψηλές τιμές σε αρκετές γειτονιές pixel του πίνακα συσχέτισης και όχι μόνο στη διαγώνιο. Αντίθετα, για την περίπτωση των MFCC χαρακτηριστικών, υπάρχουν υψηλές τιμές κυρίως στη διαγώνιο, γεγονός που υποδεικνύει χαμηλή συσχέτιση. Για να έχουμε υψηλότερη διακριτική ικανότητα επιλέγουμε τα χαρακτηριστικά με τη χαμηλότερη συσχέτιση, δηλαδή τα MFCC χαρακτηριστικά αφού αυτά είναι που θα μας δώσουν περισσότερη πληροφορία.

Βήμα 5: Εξαγωγή μοναδικού διανύσματος χαρακτηριστικών για κάθε εκφώνηση

Για την αναγνώριση των ψηφίων απαιτείται να εξαχθεί ένα διάνυσμα χαρακτηριστικών για κάθε εκφώνηση. Αυτό το διάνυσμα περιέχει τα χαρακτηριστικά mfcc, τα deltas και τα delta-deltas. Συγκεκριμένα έγινε υπολογισμός της μέση τιμής και της διασποράς όλων των παραθύρων κάθε εκφωνήσεις. Έτσι καταλήγουμε σε ένα διάνυσμα της μορφής:

$$features = \left[\mu_{mfcc}|\mu_{delta}|\mu_{delta-delta}|\sigma_{mfcc}|\sigma_{delta}|\sigma_{delta-delta}\right]_{1\times 78}$$

για κάθε εκφώνηση, συνολικά δηλαδή 133 διανύσματα. Στην εικόνα 4 φαίνεται ένα scatter plot με τις πρώτες δύο διαστάσεις του διανύσματος features.

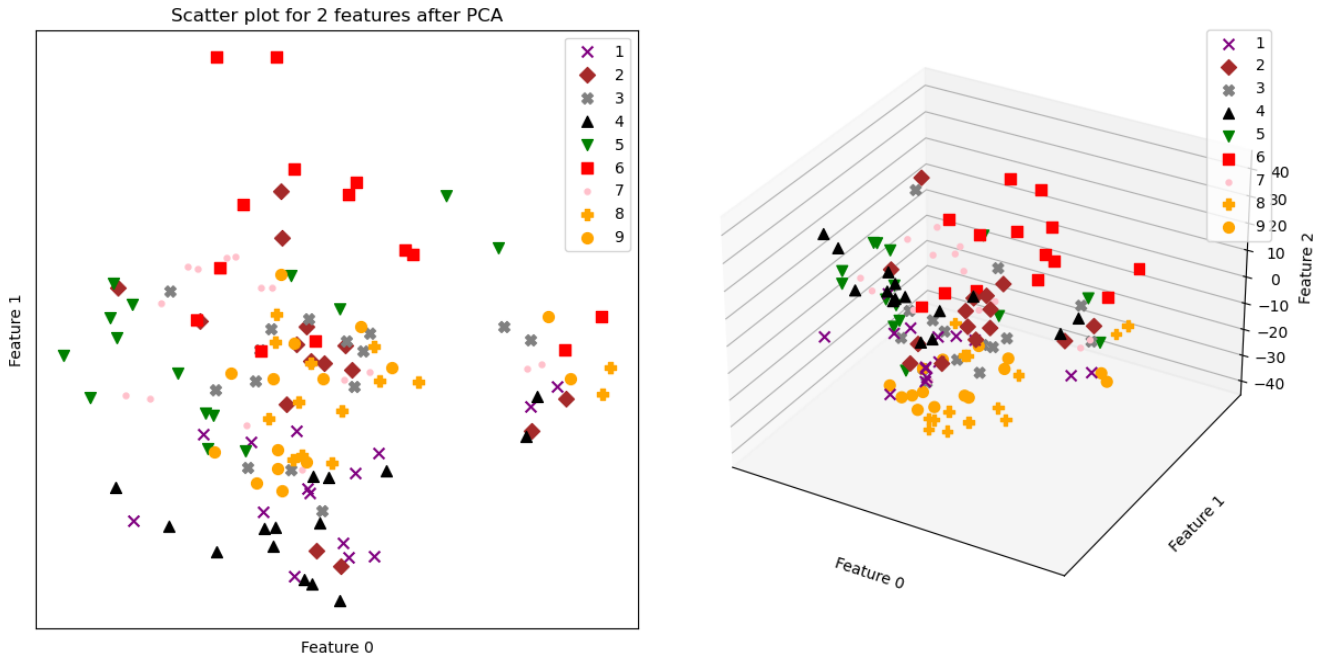


Εικόνα 4: Scatter plot για δύο χαρακτηριστικά πριν το PCA

Παρατηρούμε πως τα δείγματά μας δεν ομαδοποιούνται ιδιαιτέρως λαμβάνοντας υπόψιν μόνο τα πρώτα δύο χαρακτηριστικά και έτσι δεν μπορούμε να ορίσουμε «περιοχές» για το κάθε ψηφίο. Αυτό είναι λογικό αφού η επιλογή των δύο πρώτων χαρακτηριστικών από κάθε δείγμα έγινε αυθαίρετα.

Βήμα 6: Principal Component Analysis (PCA)

Στη συνέχεια για τη μείωση των διαστάσεων των χαρακτηριστικών εφαρμόζουμε PCA και μειώνουμε σε 2 και 3 διαστάσεις. Στην εικόνα 5 μπορούμε να δούμε ένα scatter plot για 2 και για 3 διαστάσεις.



(α) (β)

Εικόνα 5: (α) Scatter plot μετά από μείωση σε 2 διαστάσεις με PCA, (β) Scatter Plot μετά από μείωση σε 3 διαστάσεις με PCA

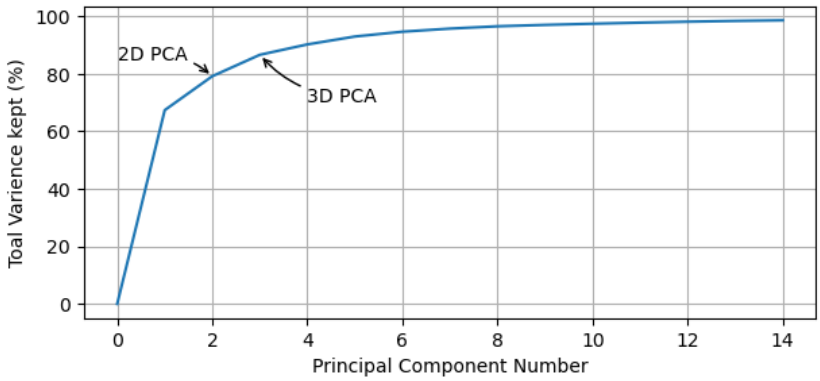
Παρατηρούμε πως με μείωση τόσο σε 2 διαστάσεις όσο και σε 3 δεν έχουμε ευκρινώς διαχωρίσιμες κλάσεις, αφού τα δείγματα μας είναι αρκετά μπλεγμένα μεταξύ τους.

Χρησιμοποιώντας .explained_variance_ratio_ μπορούμε να δούμε το ποσοστό της διασποράς των χαρακτηριστικών που κρατά η μείωση PCA. Τα στατιστικά φαίνονται στον πίνακα 3.

Διαστάσεις PCA	1η συνιστώσα	2η συνιστώσα	3η συνιστώσα	Συνολικά
2	58.79%	11.86%	-	79.18%
3	58.79%	11.86%	10.84%	86.63%

Πίνακας 3: Ποσοστό διατήρησης αρχικής διασποράς μετά την εφαρμογή PCA

Παρατηρούμε πως η μείωση σε 2 ή 3 διαστάσεις δεν είναι ικανοποιητική αφού ένα μεγάλο μέρος της αρχικής διασποράς χάνεται. Στην εικόνα 6 παρατηρούμε πως για PCA σε περίπου 7 διαστάσεις και πάνω κρατιέται ικανοποιητικό ποσοστό της διασποράς των χαρακτηριστικών.



Εικόνα 6: Σχέση αριθμού διαστάσεων PCA και κρατούμενου ποσοστού διασποράς χαρακτηριστικών

Βήμα 7: Ταξινόμηση

Προκειμένου να γίνει η ταξινόμηση των δεδομένων αρχικά τα χωρίζουμε σε train και test set με αναλογία $70\% - 30\%$ και τα κανονικοποιούμε κατά τις υποδείξεις της εκφώνησης. Για την ταξινόμηση χρησιμοποιήθηκαν 4 ταξινομητές. Τα αποτελέσματα της ταξινόμησης συμπυκνώνονται στον πίνακα 4.

Classifier	Normalization	Score
Custom Naive Bayes	No	62.5%
	Yes	57.5%
sklearn Naive Bayes	No	62.5%
	Yes	57.5%
kNeighbors ($k = 3$)	No	52.5%
	Yes	52.5%

Logistic Regression	No	72.5%
	Yes	17.5%
SVM with linear kernel	No	**75%**
	Yes	17.5%

Πίνακας 4: Ποσοστά επιτυχίας ταξινομητών

Παρατηρήσεις:

- Για όλους τους ταξινομητές παρατηρούμε καλύτερη επίδοση για μη κανονικοποιημένα δεδομένα. Αυτό είναι λογικό αφού μετά την κανονικοποίηση των δεδομένων έχουμε χάσει ένα μικρό κομμάτι πληροφορίας χρήσιμο για την ταξινόμηση.
- Την καλύτερη επίδοση πετυχαίνει ο SVM χωρίς κανονικοποιημένα δεδομένα ενώ την χειρότερη ο SVM και ο Logistic Regression με κανονικοποιημένα δεδομένα.
- Για διαφορετικά split οι επίδοσης μεταβάλλονται, αφού υπάρχει τυχαιότητα κατά τον διαχωρισμό.