Μηχανική Μάθηση

Μιχάλης Τίτσιας

Διάλεξη 7ή

Μέθοδοι ομαδοποίησης (*K-means*, μίξεις *Gaussian* κατανομών και μάθηση με τον αλγόριθμο EM)

Περιεχόμενα

Μάθηση χωρίς επίβλεψη

Ως τώρα έχουμε ασχοληθεί με μάθηση με επίβλεψη

- Τα δεδομένα εκπαίδευσης αποτελούν ζεύγη $\{\mathbf{x}_n, t_n\}_{n=1}^{N}$ και ο σκοπός μας είναι να μάθουμε μια συνάρτηση της μορφής

$$\mathbf{x} \rightarrow t$$

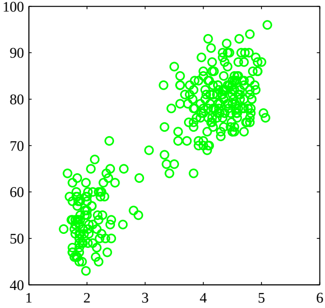
  - Αν t παίρνει διακριτές τιμές $\Rightarrow$ κατηγοριοποίηση
  - Αν t παίρνει συνεχείς τιμές $\Rightarrow$ παλινδρόμηση

Στην συνέχεια του μαθήματος θα μας απασχολήσουν τεχνικές μάθησης χωρίς επίβλεψη

- Τα δεδομένα αποτελούν μόνο μεταβλητές εισόδου $\{\mathbf{x}_n\}_{n=1}^{N}$ $\Rightarrow$ μας ενδιαφέρει να βρούμε δομή

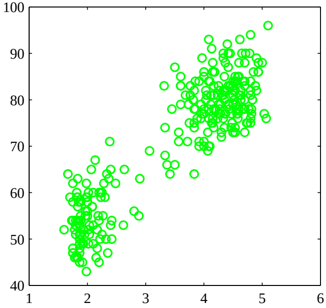
Παρακάτω θα μας απασχολήσει το πρόβλημα ομαδοποίησης (clustering), δηλ. εύρεσης ομάδων στα δεδομένα

Το πρόβλημα της ομαδοποίησης δεδομένων

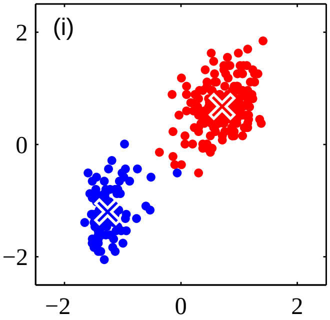


Ομαδοποίηση (clustering): Εύρεση ομάδων ώστε τα δεδομένα εντός μιας ομάδας είναι πιο όμοια απ΄ ό,τι τα δεδομένα μεταξύ διαφορετικών ομάδων

Το πρόβλημα της ομαδοποίησης δεδομένων



⇒



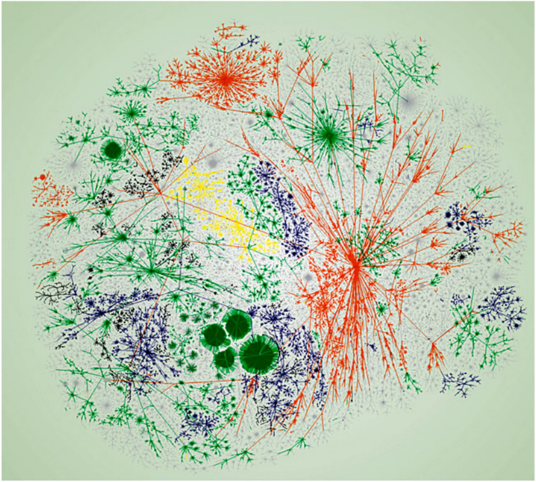
Ομαδοποίηση (clustering): Εύρεση ομάδων ώστε τα δεδομένα εντός μιας ομάδας είναι πιο όμοια απ΄ ό,τι τα δεδομένα μεταξύ διαφορετικών ομάδων

Το πρόβλημα της ομαδοποίησης δεδομένων

Πολλές εφαρμογές: Π.χ. κατάτμηση (segmentation) εικόνας

Το πρόβλημα της ομαδοποίησης δεδομένων

Πολλές εφαρμογές: Π.χ. ομαδοποίηση κόμβων σε γράφους, όπως social networks



Αλγόριθμος K-means

Δεδομένα: $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ όπου $\mathbf{x}_n \in \mathbb{R}^D$

Υποθέτουμε K ομάδες ($K \ll N$) όπου η κάθε μια συνοψίζεται από ένα μέσο διάνυσμα $\boldsymbol{\mu}_k \in \mathbb{R}^D$

Στόχος μας είναι να βρούμε ένα τρόπο να χωρίσουμε τα δεδομένα σε ομάδες βρίσκοντας κατάλληλες τιμές για τα μέσα $\boldsymbol{\mu}_k, k = 1, \ldots, K$

Αλγόριθμος K-means

Βήματα του αλγορίθμου

Αρχικοποίηση των μέσων $\boldsymbol{\mu}_k$, $k = 1, \ldots, K$

1. Ανάθεση του κάθε δεδομένου $\mathbf{x}_n$ στην ομάδα του κοντινότερου μέσου
2. Αλλαγή της τιμής του κάθε μέσου $\boldsymbol{\mu}_k$ ώστε η νέα τιμή είναι ο μέσος όρος όλων των δεδομένων που έχουν ανατεθεί στην ομάδα k (από το βήμα 2)
3. Πήγαινε στο βήμα 1 η τερμάτισε

Αλγόριθμος K-means

Βήματα του αλγορίθμου σε μορφή ψευδοκώδικα

Αρχικοποίηση των μέσων $\boldsymbol{\mu}_k$, $k = 1, \dots, K$

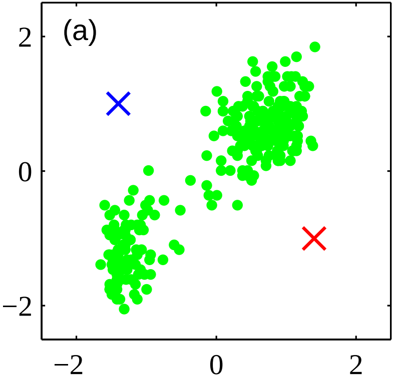
Εφόσον (Κριτήριο σύγκλισης δεν ικανοποιείται) {

1. Από $n = 1$ ως N
   $r_n =$ ομάδα (δείκτης με τιμές από 1 έως K) με την μικρότερη απόσταση $||\mathbf{x}_n - \boldsymbol{\mu}_k||^2$ από κάθε μέσο $k = 1, \dots, K$
2. Από $k = 1$ ως K
   $\boldsymbol{\mu}_k =$ μέσος όρος όλων των δεδομένων $\{\mathbf{x}_n | r_n = k\}$
3. Έλεγχος για το αν το κριτήριο σύγκλισης έχει ικανοποιηθεί

}

Οι τιμές των μεταβλητών $r_n, n = 1, \dots, N$ ονομάζονται αναθέσεις

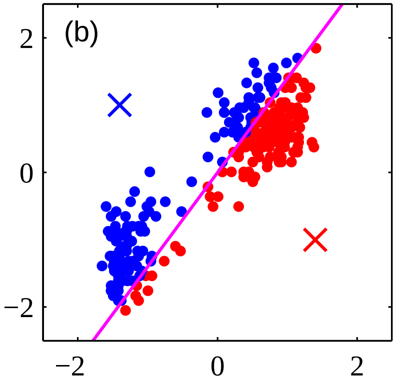
Παράδειγμα ομαδοποίησης σε δύο ομάδες



Αρχικοποίηση των μέσων

Αλγόριθμος K-means

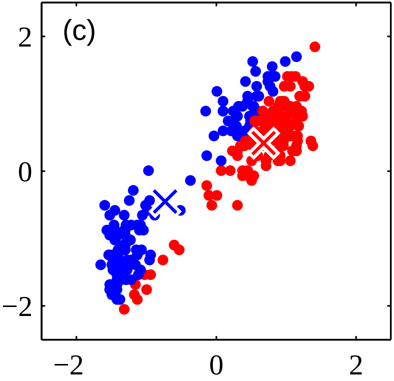
Παράδειγμα ομαδοποίησης σε δύο ομάδες



Αναθέσεις $r_n, n = 1, \ldots, N$ (επανάληψη 1)

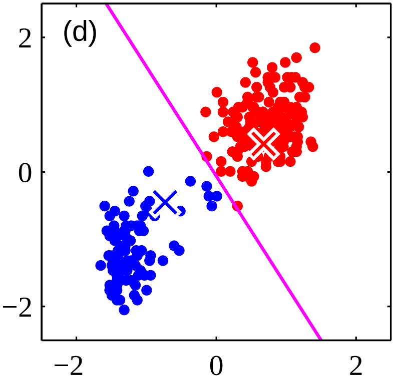
Αλγόριθμος K-means

Παράδειγμα ομαδοποίησης σε δύο ομάδες



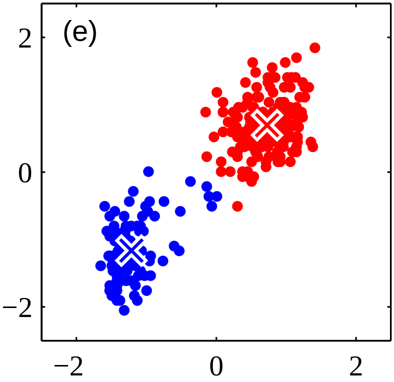
Ενημερωμένες μέσες τιμές $\boldsymbol{\mu}_k, k = 1, 2$ (επανάληψη 1)

Παράδειγμα ομαδοποίησης σε δύο ομάδες



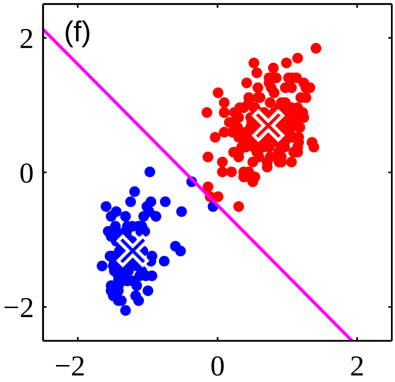
Αναθέσεις $r_n, n = 1, \ldots, N$ (επανάληψη 2)

Παράδειγμα ομαδοποίησης σε δύο ομάδες



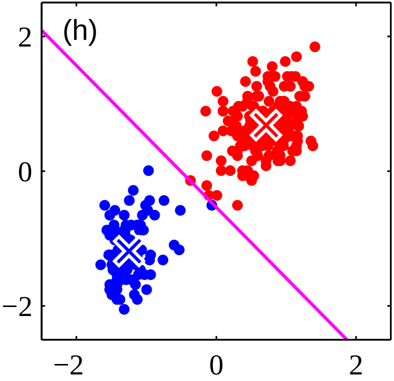
Ενημερωμένες μέσες τιμές $\boldsymbol{\mu}_k, k = 1, 2$ (επανάληψη 2)

Παράδειγμα ομαδοποίησης σε δύο ομάδες



Αναθέσεις και μέσες τιμές (επανάληψη 3)

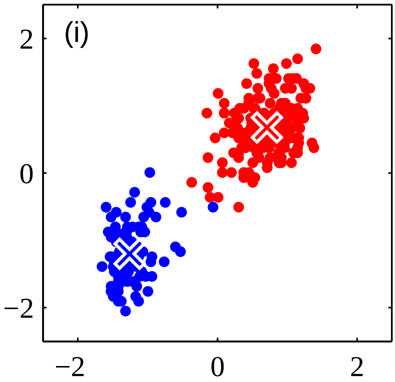
Παράδειγμα ομαδοποίησης σε δύο ομάδες



Αναθέσεις και μέσες τιμές (επανάληψη 4)

Αλγόριθμος K-means

Παράδειγμα ομαδοποίησης σε δύο ομάδες



Τελικά μέσα και ομαδοποίηση

Αλγόριθμος K-means

Ποια είναι η συνάρτηση κόστους που βελτιστοποιεί ο K-means;

Απ' ό,τι έχουμε πει ως τώρα μπορούμε να συμπεράνουμε ότι μια τέτοια συνάρτηση θα εξαρτάται από τις αναθέσεις $(r_1, \ldots, r_N)$ και τα μέσα $(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K)$. Δηλ. θα είναι της μορφής

$$J(r_1, \ldots, r_N, \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K)$$

Το να γνωρίζουμε αυτή την συνάρτηση J θα χρησίμευε στην πράξη στον έλεγχο σύγκλισης αλλά (και πιο θωρητικά) στην επιπλέον κατανόηση του αλγορίθμου

Συνάρτηση κόστους του K-means

- Ισοδύναμα αναπαριστούμαι τις αναθέσεις ως δυαδικά διανύσματα υπόδειξης, δηλ. κάθε ανάθεση είναι ένα διάνυσμα $\mathbf{r}_n = (r_{n1}, \ldots, r_{nK})$ τέτοιο ώστε

$$r_{nk} \in \{0, 1\}, \quad \sum_{k=1}^{K} r_{nk} = 1$$

  (όταν το $\mathbf{x}_n$ ανατίθεται στην ομάδα k, τότε $r_{nk} = 1$ και $r_{nj} = 0$ για $j \neq k$)
- Η συνάρτηση κόστους του K-means είναι

$$J(\mathbf{r}_1, \ldots, \mathbf{r}_N, \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K) = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} ||\mathbf{x}_n - \boldsymbol{\mu}_k||^2$$

Αλγόριθμος K-means

Συνάρτηση κόστους του K-means

$$J(\mathbf{r}_1, \dots, \mathbf{r}_N, \boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_K) = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} ||\mathbf{x}_n - \boldsymbol{\mu}_k||^2$$

Ο αλγόριθμος K-means ελαχιστοποιεί αυτή την συνάρτηση ως προς $(\mathbf{r}_1, \dots, \mathbf{r}_N)$ και $(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_K)$ επαναλαμβάνοντας τα βήματα

- Δοθέντος των $(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_K)$ ελαχιστοποιούμε ως προς τις αναθέσεις $(\mathbf{r}_1, \dots, \mathbf{r}_N)$
- Δοθέντος των αναθέσεων $(\mathbf{r}_1, \dots, \mathbf{r}_N)$ ελαχιστοποιούμε ως προς τα μέσα $(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_K)$

Κάθε βήμα του αλγορίθμου μειώνει (ή σε περίπτωση σύγκλισης αφήνει αμετάβλητη) την συνάρτηση κόστους J

Αλγόριθμος K-means

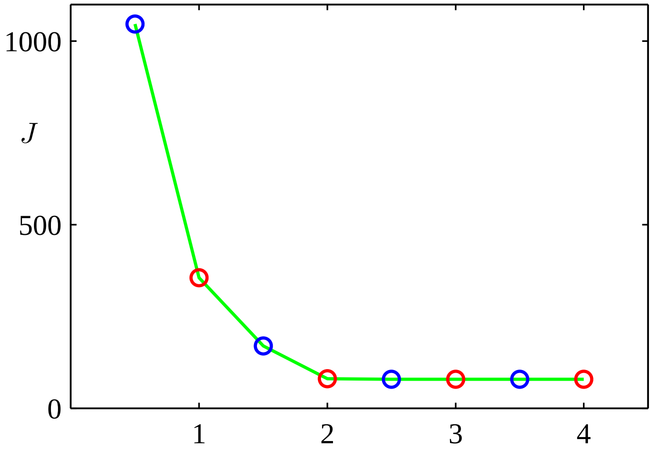
Συνάρτηση κόστους του K-means

$$J(\mathbf{r}_1, \ldots, \mathbf{r}_N, \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K) = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} ||\mathbf{x}_n - \boldsymbol{\mu}_k||^2$$

Ο αλγόριθμος K-means ελαχιστοποιεί αυτή την συνάρτηση ως προς $(\mathbf{r}_1, \ldots, \mathbf{r}_N)$ και $(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K)$ επαναλαμβάνοντας τα βήματα

- Δοθέντος των $(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K)$ ελαχιστοποιούμε ως προς τις αναθέσεις $(\mathbf{r}_1, \ldots, \mathbf{r}_N)$
  - $\Rightarrow$ ανάθεση του κάθε δεδομένου $\mathbf{x}_n$ στην ομάδα του κοντινότερου μέσου
- Δοθέντος των αναθέσεων $(\mathbf{r}_1, \ldots, \mathbf{r}_N)$ ελαχιστοποιούμε ως προς τα μέσα $(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K)$
  - $\Rightarrow \boldsymbol{\mu}_k = \frac{\sum_{n=1}^{N} r_{nk}\mathbf{x}_n}{\sum_{n=1}^{N} r_{nk}}$, $k = 1, \ldots, K$, που δεν είναι τίποτα άλλο παρά μέσοι όροι των δεδομένων της κάθε ομάδας

Αλγόριθμος K-means



Η εξέλιξη της τιμής της συνάρτησης κόστους ως προς τις επαναλήψεις (οριζόντιος άξονας) του K-means στο παράδειγμα που είδαμε νωρίτερα

Αλγόριθμος K-means

Παράδειγμα εφαρμογής: Κατάτμηση και συμπίεση εικόνας. Κάθε δεδομένο είναι ένα pixel, συγκεκριμένα ένα 3-διάστατο διάνυσμα $\mathbf{x} = [x_r \; x_g \; x_b]$ με τις τιμές χρώματος (ονομάζονται και τιμές RGB)

Original image

Θα θέλαμε να χωρίσουμε την αρχική εικόνα σε υποπεριοχές $\Rightarrow$ κατάτμηση εικόνας

Αλγόριθμος K-means

Εφαρμόζουμε K-means για διαφορετικές τιμές του K. Για συγκεκριμένο K η κατατμημένη εικόνα κατασκευάζεται αντικαθιστώντας την αρχική τιμή του pixel $\mathbf{x}_n$ με την τιμή του μέσου $\boldsymbol{\mu}_k$ της ομάδας στην οποία ανήκει

Αλγόριθμος K-means

Η ομαδοποίηση μας δίνει επίσης μια μέθοδο συμπίεσης των δεδομένων (vector quantization)

$K = 2$

$K = 3$

$K = 10$

Original image

Αλγόριθμος K-means

Επίλογη του αριθμού των ομάδων K

- Καθώς μεγαλώνει το K τόσο υπάρχει η δυνατότητα η συνάρτηση κόστους J να πάρει ολοένα και μικρότερη τιμή
  - Αν επιλέξουμε K ίσο με τον αριθμό των δεδομένων και θέσουμε $\boldsymbol{\mu}_n = \mathbf{x}_n$, τότε $J = 0$!
- Η επιλογή του K γίνεται από τον χρήστη συνήθως βάσει κάποιου κριτηρίου που έχει να κάνει με τον αρχικό σκόπο εφαρμογής του K-means
  - Π.χ. στο πρόβλημα συμπίεσης εικόνας, το K θα μπορούσε να επιλέγεται βάσει κάποιου επιθυμητού επιπέδου συμπίεσης ή επιθυμήτου ελάχιστου σφάλματος ανακατασκευής της εικόνας
- Αν ποσοτικοποιήσουμε το κριτήριο επίδοσης, τότε θα μπορούσαμε να επιλέξουμε το K με cross-validation

Αλγόριθμος K-means

Μειονεκτήματα του K-means

- Hard assingment των δεδομένων σε ομάδες: Είτε το δεδομένο βρίσκεται κοντά στο μέσο της ομάδας είτε κοντά στο σύνορο με μια διπλανή ομάδα, η ανάθεση γίνεται με την ίδια βεβαιότητα!
  - Μια καλύτερη λύση θα ήταν η ανάθεση να γίνεται με αβεβαιότητα, δηλ. πιθανοτικά
- Υποθέτει ότι οι ομάδες έχουν σφαιρικό σχήμα

Η πιθανοτική ομαδοποίηση με μίξεις Gaussian κατανομών βελτιώνει σημαντικά τα παραπάνω

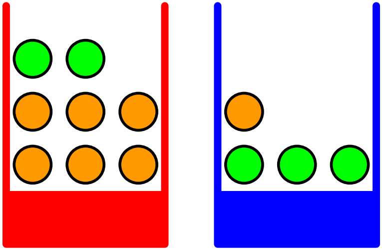
Πιθανοτική ομαδοποίηση

Η πιθανοτική ομαδοποίηση θα βασιστεί στην εισαγωγή ένος πιθανοτικού μοντέλου που στοχεύει να περιγράψει τον μηχανισμό γέννησης των δεδομένων σε ένα πρόβλημα ομαδοποίησης

Η όλη ιδέα είναι ουσιαστικά όμοια με το παράδειγμα των «κουτιών με τα φρούτα» (Διάλεξη 2ή) το οποίο αξίζει να θυμηθούμε υπό ένα νέο πρίσμα

Πιθανοτική ομαδοποίηση

Πορτοκάλια και μήλα ως πιθανοτικό μοντέλο ομαδοποίησης



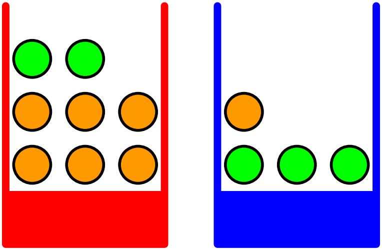
Παραγωγή δεδομένων ως εξής:

1. Επέλεξε ένα από δύο κουτιά με εκ των προτέρων πιθανότητες π_1 και π_2
2. Δοθέντος του κουτιού που επιλέχθηκε στο 1), επέλεξε ένα φρούτο/δεδομένο

Παρατηρούμε ότι ο ρόλος τους κουτιού «μοιάζει» με την έννοια της ομάδας

Πιθανοτική ομαδοποίηση

Πορτοκάλια και μήλα ως πιθανοτικό μοντέλο ομαδοποίησης



Το πιθανοτικό μοντέλο για την παραγωγή κάθε δεδομένου περιγράφεται από δύο τυχαίες μεταβλητές:

1. z : Επιλέγεται με βάσει τις πιθανότητες π_1 και π_2 και ταυτοποιεί το κουτί/ομάδα της οποίας μέλος θα είναι το δεδομένο που θα παραχθεί
2. $\mathbf{x}$: Το δεδομένο που επιλέγεται δοθέντος της ομάδας που έχει καθοριστεί από το z, δηλ. το $\mathbf{x}$ παράγεται από μια υπό συνθήκη κατανομή
$$p(\mathbf{x}|z)$$

Πιθανοτική ομαδοποίηση

Το πιθανοτικό μοντέλο ομαδοποίησης (για K ομάδες) υποθέτει ότι κάθε δεδομένο παράγεται ως εξής

- Καθορισμός της ομάδας από την οποία θα παραχθεί το δεδομένο $\Rightarrow$ καθορισμός της τιμής μιας τυχαίας μεταβλητής $\mathbf{z}$

$$p(\mathbf{z}) = \prod_{k=1}^{K} \pi_k^{z_k}$$

  όπου π_k είναι η εκ των προτέρων πιθανότητα της ομάδας k

- Παραγωγή του δεδομένου δοθέντος του $\mathbf{z}$

$$p(\mathbf{x}|\mathbf{z}) = \prod_{k=1}^{K} [p(\mathbf{x}|k)]^{z_k}$$

  όπου $p(\mathbf{x}|k)$ είναι μια υπό συνθήκη κατανομή που περιγράφει το πως παράγονται τα δεδομένα της ομάδας k

Πιθανοτική ομαδοποίηση

- Το $\mathbf{z}$, που ταυτοποιεί την ομάδα του δεδομένου $\mathbf{x}$, στην πράξη δεν θα είναι γνωστό!
- Υπό αυτή την έννοια το $\mathbf{z}$ είναι αυτό που αναφέρεται ως κρυμμένη μεταβλητή
- Μόνο το δεδομένο $\mathbf{x}$ παρατηρείται και είναι γνωστό
- Η ολική πιθανότητα/κατανομή του $\mathbf{x}$ υπολογίζεται εφαρμοζόντας τον κανόνα αθροίσματος και περιθωριοποιώντας το άγνωστο $\mathbf{z}$

$$p(\mathbf{x}) = \sum_{\mathbf{z}} p(\mathbf{z})p(\mathbf{x}|\mathbf{z}) = \sum_{k=1}^{K} \pi_k p(\mathbf{x}|k)$$

Η κατανομή $p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k p(\mathbf{x}|k)$ ονομάζεται μίξη κατανομών

- $\Rightarrow$ με μια εξίσωση η μίξη κατανομών περιγράφει όλο το πιθανοτικό μοντέλο για ομαδοποίηση δεδομένων

Μίξεις Gaussian κατανομών

- **Μίξη κατανομών**:

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k p(\mathbf{x}|k), \quad 0 \leq \pi_k \leq 1, \quad \sum_{k=1}^{K} \pi_k = 1$$

- Η υπό συνθήκη κατανομή $p(\mathbf{x}|k)$ ονομάζεται συνιστώσα κατανομή (mixture component)
- Η εκ των προτέρων πιθανότητα της ομάδας π_k ονομάζεται συντελεστής μίξης (mixing coefficient)

Συγκεκριμένες περιπτώσεις αυτών των μοντέλων προκύπτουν επιλέγοντας συγκεκριμένη συναρτησιακή μορφή για τις συνιστώσες κατανομές $p(\mathbf{x}|k)$

- Η συνηθέστερη είναι να επιλέξουμε για κάθε $p(\mathbf{x}|k)$ μια Gaussian

Μίξεις Gaussian κατανομών

- Η Gaussian κατανομή

$$p(\mathbf{x}|k) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \Sigma_k) = \frac{1}{(2\pi)^{\frac{D}{2}}|\Sigma_k|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_k)^T \Sigma_k^{-1}(\mathbf{x}-\boldsymbol{\mu}_k)}$$

  έχει μόνο μια κορυφή $\Rightarrow$ είναι ένα πιθανοτικό μοντέλο για μια μόνο ομάδα δεδομένων!
- Ορισμός: **Μίξη Gaussian κατανομών**:

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \Sigma_k)$$

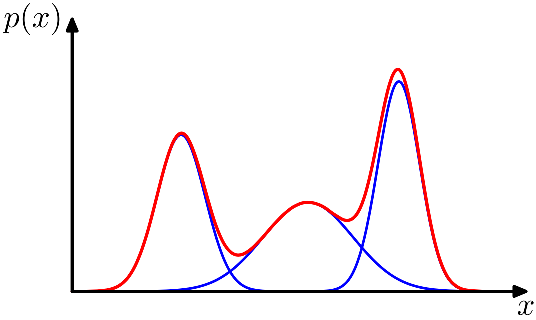
$$0 \le \pi_k \le 1, \quad \sum_{k=1}^{K} \pi_k = 1$$

- Αποτελεί ένα μοντέλο για πολλές ομάδες δεδομένων όπου η κάθε ομάδα περιγράφεται από μια Gaussian κατανομή

Μίξεις Gaussian κατανομών

- **Μίξη Gaussian κατανομών**:

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \Sigma_k), \quad 0 \leq \pi_k \leq 1, \quad \sum_{k=1}^{K} \pi_k = 1$$

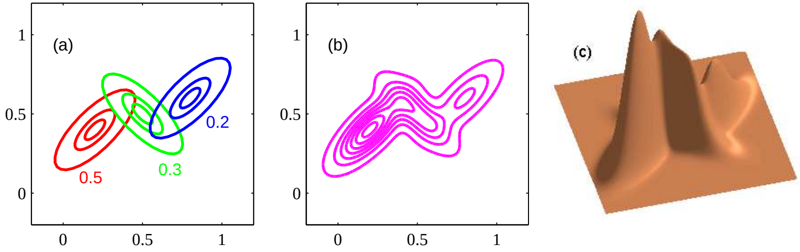


Μίξη Gaussian κατανομών (κόκκινη γραμμή) για μονοδιάστατα δεδομένα. Οι μπλε γραμμές δείχνουν τους όρους $\pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \Sigma_k)$

Μίξεις Gaussian κατανομών

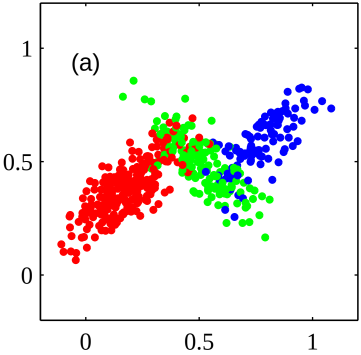
- **Μίξη Gaussian κατανομών**:

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \Sigma_k), \quad 0 \leq \pi_k \leq 1, \quad \sum_{k=1}^{K} \pi_k = 1$$



Μίξη Gaussian κατανομών (ροζ γραμμή) για δισδιάστατα δεδομένα. Οι ελλείψεις δείχνουν τους όρους $\pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \Sigma_k)$

Μίξεις Gaussian κατανομών



Μίξη από τρεις ($K = 3$) Gaussian θα μπορούσε να είχε παράγει τα δεδομένα του σχήματος τα οποία ανήκουν σε τρεις ομάδες

Η θέση και το ελλειψοειδές σχήμα της κάθε ομάδας k καθορίζεται πλήρως από τις τιμές των παραμέτρων $(\boldsymbol{\mu}_k, \Sigma_k)$

Η συχνότητα εμφάνισης δεδομένων μιας ομάδας εξαρτάται από την παράμετρο π_k

Μίξεις Gaussian κατανομών

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \Sigma_k), \quad 0 \leq \pi_k \leq 1, \quad \sum_{k=1}^{K} \pi_k = 1$$

- Οι παραμέτροι του μοντέλου είναι

$$\{\pi_k, \boldsymbol{\mu}_k, \Sigma_k\}_{k=1}^{K}$$

  τις οποίες θα θέλαμε να εκτιμήσουμε μέσω ενός αλγορίθμου εκπαίδευσης (θα μιλήσουμε σε λίγο για αυτό)
- Έστω ότι έχουμε εκτιμήσει τις παραπάνω παραμέτρους
- Πώς θα εξάγαμε κάποιο συμπέρασμα για την άγνωστη ομάδα ενός δεδομένου $\mathbf{x}$;

Μίξεις Gaussian κατανομών

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \Sigma_k), \quad 0 \leq \pi_k \leq 1, \quad \sum_{k=1}^{K} \pi_k = 1$$

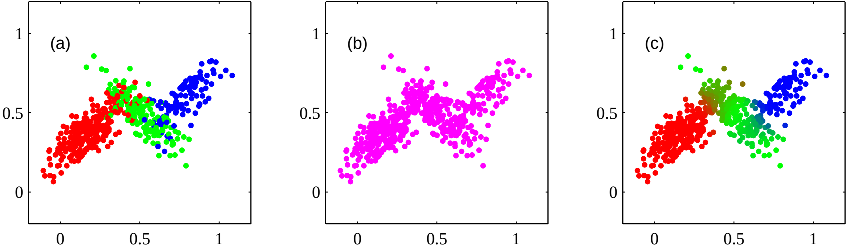
- Οι παραμέτροι του μοντέλου είναι

$$\{\pi_k, \boldsymbol{\mu}_k, \Sigma_k\}_{k=1}^{K}$$

  τις οποίες θα θέλαμε να εκτιμήσουμε μέσω ενός αλγορίθμου εκπαίδευσης (θα μιλήσουμε σε λίγο για αυτό)
- Έστω ότι έχουμε εκτιμήσει τις παραπάνω παραμέτρους
- Πώς θα εξάγαμε κάποιο συμπέρασμα για την άγνωστη ομάδα ενός δεδομένου $\mathbf{x}$;
  - $\Rightarrow$ εκ των υστέρων πιθανότητα (θεώρημα Bayes):

$$\gamma(z_k) = p(z_k = 1|\mathbf{x}) = \frac{\pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_j, \Sigma_j)}$$

Μίξεις Gaussian κατανομών



(a) (b) (c)

- (a): Πως έχουν παραχθεί στη πραγματικότητα τα δεδομένα (το οποίο είναι άγνωστο στην πράξη)
- (b): Τί παρατηρούμαι
- (c): Εκ των υστέρων πιθανότητες

Μίξεις Gaussian κατανομών

Μέγιστη πιθανοφάνεια: Εκτίμηση των παραμέτρων $\{\pi_k, \boldsymbol{\mu}_k, \Sigma_k\}_{k=1}^{K}$

- Δοθέντος ενός συνόλου δεδομένων εκπαίδευσης $(\mathbf{x}_1, \ldots, \mathbf{x}_N)$, υποθέτουμε ότι το καθένα έχει παραχθεί ανεξάρτητα από τη μίξη $p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \Sigma_k)$
- Από κοινού κατανομή

$$p(\mathbf{X}|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \prod_{n=1}^{N} \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \Sigma_k)$$

- Λογαριθμική πιθανοφάνεια

$$\mathcal{L}(\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{n=1}^{N} \log \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \Sigma_k)$$

- Ως συνήθως θέλουμε να την μεγιστοποιήσουμε ως προς τις παραμέτρους
  - $\Rightarrow$ δύσκολο λόγω του αθροίσματος που υπάρχει μέσα στον λογάριθμο

Αλγόριθμος EM για μίξεις Gaussian κατανομών

$$\mathcal{L}(\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{n=1}^{N} \log \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_k, \Sigma_k)$$

- Η μεγιστοποίηση γίνεται με τον αλγόριθμο **EM (expectation-maximization)**
  - Ένας από τους σημαντικότερους αλγορίθμους στη μηχανική μάθηση και εφαρμοσμένη στατιστική
  - Εφαρμόζεται σε πάρα πολλά μοντέλα $\Rightarrow$ όχι μόνο σε μίξεις από Gaussian
- Θα παράγουμε πρώτα τις εξισώσεις του EM για μίξεις Gaussian κατανομών με «εμπειρικό» τρόπο και έπειτα (επόμενο μάθημα) θα περιγράψουμε τον αλγόριθμο στην γενική του μορφή που θα περιλαμβάνει την χρήση του EM και για άλλα μοντέλα

Αλγόριθμος EM για μίξεις Gaussian κατανομών

$$\mathcal{L}(\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{n=1}^{N} \log \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_k, \Sigma_k)$$

- Αφού θέλουμε να μεγιστοποίησουμε την $\mathcal{L}$ παίρνουμε παραγώγους και θέτουμε στο μηδέν. Έπειτα από πολλές πράξεις καταλήγουμε στις εξισώσεις

$$\boldsymbol{\mu}_k = \frac{\sum_{n=1}^{N} \gamma(z_{nk}) \mathbf{x}_n}{\sum_{n=1} \gamma(z_{nk})}$$

$$\Sigma_k = \frac{\sum_{n=1}^{N} \gamma(z_{nk})(\mathbf{x}_n - \boldsymbol{\mu}_k)(\mathbf{x}_n - \boldsymbol{\mu}_k)^T}{\sum_{n=1}^{N} \gamma(z_{nk})}$$

$$\pi_k = \frac{\sum_{n=1}^{N} \gamma(z_{nk})}{N}$$

- Οι εξισώσεις αυτές δεν αποτελούν αναλυτικές λύσεις, αφού οι παράμετροι εμφανίζονται και στις δύο πλευρές της εξίσωσης (π.χ. εντός του κάθε $\gamma(z_{nk})$)
  - ωστόσο μας δίνουν την ιδέα για έναν επαναληπτικό αλγόριθμο

Αλγόριθμος EM για μίξεις Gaussian κατανομών

Αλγόριθμος EM

1. Αρχικοποίηση παραμέτρων $\{\pi_k, \boldsymbol{\mu}_k, \Sigma_k\}_{k=1}^{K}$
2. **E βήμα:**

$$\gamma(z_{nk}) = \frac{\pi_k \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_j, \Sigma_j)}, \quad n = 1, \ldots, N, \quad k = 1, \ldots, K$$

3. **M βήμα:**

$$\boldsymbol{\mu}_k^{new} = \frac{\sum_{n=1}^{N} \gamma(z_{nk}) \mathbf{x}_n}{\sum_{n=1} \gamma(z_{nk})}, \quad k = 1, \ldots, K$$

$$\Sigma_k^{new} = \frac{\sum_{n=1}^{N} \gamma(z_{nk})(\mathbf{x}_n - \boldsymbol{\mu}_k^{new})(\mathbf{x}_n - \boldsymbol{\mu}_k^{new})^T}{\sum_{n=1}^{N} \gamma(z_{nk})}, \quad k = 1, \ldots, K$$

$$\pi_k^{new} = \frac{\sum_{n=1}^{N} \gamma(z_{nk})}{N}, \quad k = 1, \ldots, K$$

4. Υπολόγισε την τρέχουσα τιμή της λογαριθμικής πιθανοφάνειας και έπειτα από έλεγχο σύγκλισης, πήγαινε στο 2 ή τερμάτισε

Αλγόριθμος EM για μίξεις Gaussian κατανομών

Πρόσεξε ότι υπάρχει κάποια ομοιότητα με τον K-means

Τα δύο βασικά βήματα του K-means ήταν

1. Εύρεση των αναθέσων r_{nk}
2. Ενημέρωση των μέσων

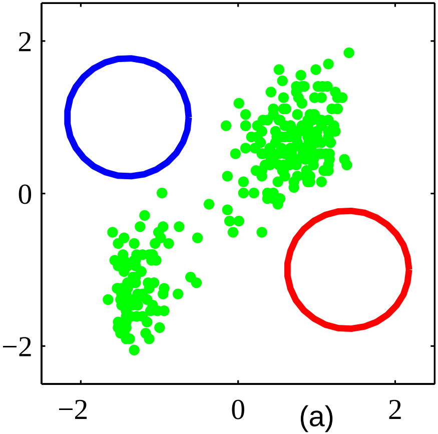
$$\boldsymbol{\mu}_k = \frac{\sum_{n=1}^{N} r_{nk}\mathbf{x}_n}{\sum_{n=1}^{N} r_{nk}}, \quad k = 1, \ldots, K$$

Ο EM έχει κάποια αντίστοιχα βήματα

1. Εύρεση πιθανοτικών αναθέσεων $\gamma(z_{nk}) = \frac{\pi_k \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_j, \Sigma_j)}$
2. Ενημέρωση των παραμέτρων και όπου η σχέση ενημέρωσης για τα μέσα είναι όμοια με αυτή του K-means

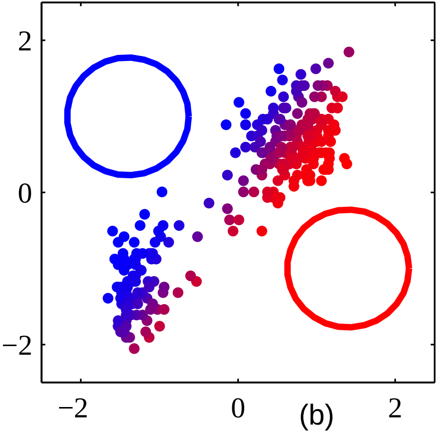
$$\boldsymbol{\mu}_k = \frac{\sum_{n=1}^{N} \gamma(z_{nk})\mathbf{x}_n}{\sum_{n=1} \gamma(z_{nk})}, \quad k = 1, \ldots, K$$

Παράδειγμα



Αρχικοποίηση των Gaussian υπό συνθήκη κατανομών

Παράδειγμα

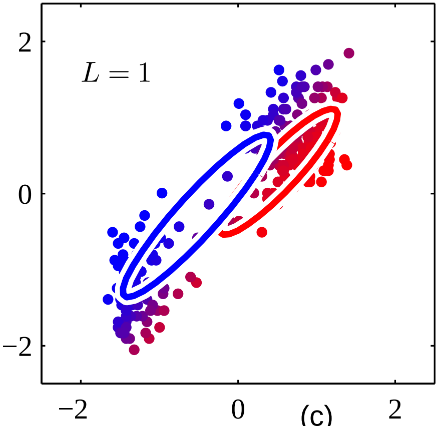


(b)

Υπολογισμός των πιθανοτικών αναθέσων $\gamma(z_{nk})$

Αλγόριθμος EM για μίξεις Gaussian κατανομών

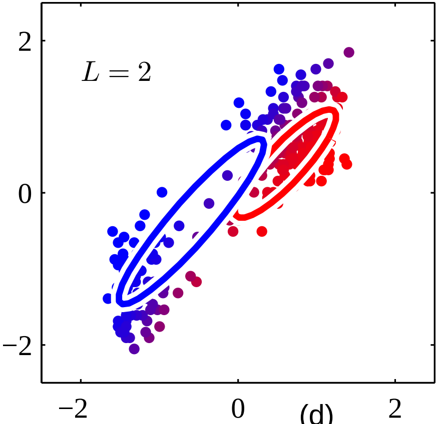
Παράδειγμα



(c)

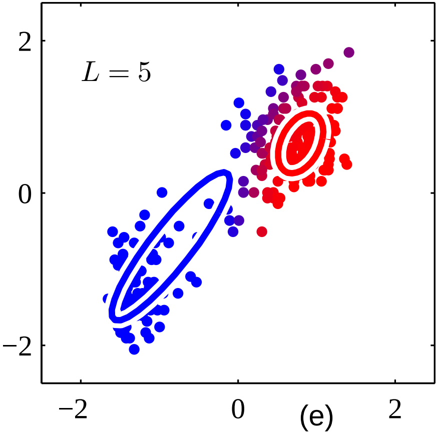
Οι μορφή των Gaussian υπό συνθήκη κατανομών μετά την ενημέρωση στην πρώτη επανάληψη

Παράδειγμα



(d)

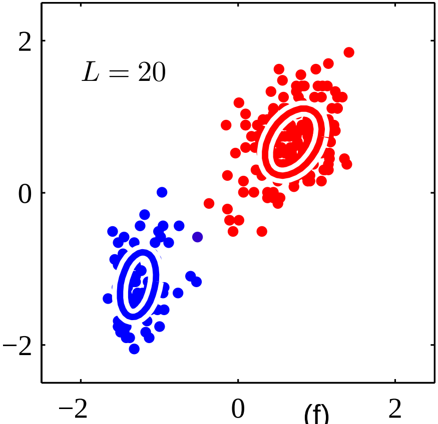
Παράδειγμα



(e)

Αλγόριθμος EM για μίξεις Gaussian κατανομών

Παράδειγμα



(f)

Σχέση EM με K-means

- Έστω ότι εφαρμόζουμε τον EM για μίξεις από Gaussian όπου οι συνιστώσες κατανομές έχουν τη μορφή

$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \Sigma_k) = \frac{1}{(2\pi\epsilon)^{\frac{D}{2}}} e^{-\frac{1}{2\epsilon}||\mathbf{x}-\boldsymbol{\mu}_k||^2}$$

όπου $\Sigma_k = \epsilon I$ και η παράμετρος ϵ είναι σταθερή $\Rightarrow$ δεν εκπαιδεύεται μέσω EM

- Τότε οι εκ των υστέρων πιθανότητες

$$\gamma(z_{nk}) = \frac{\pi_k e^{-\frac{1}{2\epsilon}||\mathbf{x}_n-\boldsymbol{\mu}_k||^2}}{\sum_{j=1}^{K} \pi_j e^{-\frac{1}{2\epsilon}||\mathbf{x}_n-\boldsymbol{\mu}_j||^2}}$$

γίνονται οι hard assingments του K-means (δηλ. $\gamma(z_{nk}) \rightarrow r_{nk}$) καθώς $\epsilon \rightarrow 0$

- Οπότε ο K-means μπορεί να ληφθεί ως ειδική περίπτωση του EM

Επίλογος

- Διάβασμα για το σπίτι: Bishop: κεφάλαιο 9 από την αρχή ώς τη σελίδα 439 και επίσης ενότητα 9.3.2
- Επόμενο μάθημα: EM αλγόριθμος γενικά, ομαδοποίηση δυαδικών δεδομένων με μίξεις από Bernulli κατανομές