

Τεχνητή Νοημοσύνη

15η διάλεξη (2023-24)

Ίων Ανδρουτσόπουλος

<http://www.aueb.gr/users/ion/>

Οι διαφάνειες αυτής της διάλεξης βασίζονται σε ύλη των βιβλίων (α) *Artificial Intelligence – A Modern Approach* των S. Russel και P. Norvig, 2^η και 4^η έκδοση, Prentice Hall, 2003 και 2020, (β) *Τεχνητή Νοημοσύνη των Βλαχάβα κ.ά.*, 3^η έκδοση, Β. Γκιούρδας Εκδοτική, 2006 και (γ) *Machine Learning* του T. Mitchell, McGraw-Hill, 1997. Τα περισσότερα σχήματα των διαφανειών βασίζονται σε αντίστοιχα σχήματα των διαφανειών που συνοδεύουν τα πρώτα δύο βιβλία.

Τι θα ακούσετε σήμερα

- Κατάταξη κειμένων σε κατηγορίες με χρήση μηχανικής μάθησης.
- Εντροπία και κέρδος πληροφορίας.
- Επιλογή ιδιοτήτων μέσω κέρδους πληροφορίας και μέσω αναζήτησης υποσυνόλων ιδιοτήτων.
- Αλγόριθμος των k -κοντινότερων γειτόνων.
- Μη επιβλεπόμενη ομαδοποίηση με τον k -means.

Διαφημιστικά μηνύματα

our highly successful multi – national company gives you an exclusive business that generates an extra weekly income of up to \$ 600 or more ... anyone can easily make money ... if you wish to be removed from our list ...

call for papers 9 th european workshop on natural language generation ... is a subfield of natural language processing that generates texts in human languages from non-linguistic data or knowledge ... for the systems to be successful ...

- Πρόβλημα **κατάταξης κειμένων**
 - Εδώ 2 κατηγορίες: **ανεπιθύμητα** ($C = 1$), **επιθυμητά** ($C = 0$).
 - Αναζητούμε μια συνάρτηση που να κατατάσσει κάθε κείμενο στη σωστή κατηγορία.

Εξόρυξη γνώμης

The screenshot shows a search interface for 'Apple Mac mini'. At the top, there is a search bar containing the text 'Apple Mac mini' and a blue 'GO' button. Below the search bar, there are four rows of sentiment analysis results, each represented by a white rounded rectangle with a vertical line separating the text from the star rating. The first row is for 'money, price, cost, ...' with a 5-star rating. The second row is for 'ram, memory, ...' with a 3-star rating. The third row is for 'design, color, feeling, ...' with a 4-star rating. The fourth row is for 'extras, keyboard, screen, ...' with a 2-star rating.

Category	Rating (Stars)
money, price, cost, ...	5
ram, memory, ...	3
design, color, feeling, ...	4
extras, keyboard, screen, ...	2

- Από **κριτικές πελατών** (<http://alt.qcri.org/semEval2016/task5/>).
- Από **tweets** (<http://alt.qcri.org/semEval2016/task4/>), **FB updates**, **blogs**, **εφημερίδες**, **ραδιοφωνικές/τηλεοπτικές εκπομπές**, ...

Κατάταξη κειμένων

- Στη γενικότερη περίπτωση n κατηγορίες.
 - Π.χ. οικονομικές ειδήσεις, αθλητικές, κοινωνικές κλπ.
 - Π.χ. τοποθεσία, τιμές, μενού, εξυπηρέτηση κ.λπ. εστιατορίου.
 - Π.χ. θετική, αρνητική, μικτή, ουδέτερη γνώμη.
- Προσέγγιση με **χειρωνακτική** κατασκευή κανόνων:
 - Π.χ. αν περιέχει τη λέξη «money» και στις 5 τελευταίες γραμμές τη λέξη «remove» και ... τότε είναι spam.
 - **Δυσκολία** χειρωνακτικής κατασκευής κανόνων που να επιτυγχάνουν **υψηλά ποσοστά ορθότητας**.
 - **Δυσκολία προσαρμογής** στις ιδιαιτερότητες και τα ενδιαφέροντα των χρηστών.
- Προσέγγιση με επιβλεπόμενη (ή ημι-επιβλεπόμενη) **μηχανική μάθηση**:
 - Στην πράξη συχνά επιτυγχάνει **καλύτερα αποτελέσματα**.
 - Ευκολία προσαρμογής και **επανεκπαίδευσης**.

Κατάταξη κειμένων με επιβλεπόμενη MM

παραδείγματα **επιθυμητών** και **ανεπιθύμητων** μηνυμάτων + σωστές απαντήσεις

εκπαίδευση

προεπεξεργασία

μονάδα εκπαίδευσης

συνάρτηση ταξινόμησης

προεπεξεργασία

μονάδα ταξινόμησης

νεοεισερχόμενο μήνυμα

απόφαση: **επιθυμητό** ή **ανεπιθύμητο**

χρήση

Προεπεξεργασία κειμένων

our highly successful multi – national company gives you an exclusive business that generates an extra weekly income of up to \$ 600 or more ... anyone can easily make money ... if you wish to be removed from our list ...

< money:1, language:0, natural:0, \$:1, adult:0, call:0, exclusive:1, **successful:1**, removed:1, **generates:1**, ...>

call for papers 9 th european workshop on natural language generation ... is a subfield of natural language processing that generates texts in human languages from non-linguistic data or knowledge ... for the systems to be successful ...

< money:0, language:1, natural:1, \$:0, adult:0, call:1, exclusive:0, **successful:1**, removed:0, **generates:1**, ...>

- Εναλλακτικά οι ιδιότητες μπορεί να παριστάνουν **συχνότητες εμφάνισης λέξεων** στο μήνυμα και/ή να αντιστοιχούν σε **φράσεις** (π.χ. «to be removed»), **χαρακτηριστικά εικόνων** κλπ.

Επιλογή ιδιοτήτων

- Για **ποιες λέξεις** (ή φράσεις ή ...) θα υπάρχουν **ιδιότητες** στα διανύσματα;
 - 1ο βήμα: μόνο λέξεις που εμφανίζονται **τουλάχιστον** k φορές στα παραδείγματα εκπαίδευσης (π.χ. $k = 3$).
 - Συνήθως παραμένουν αρκετές **χιλιάδες** λέξεις (και πολύ περισσότερες φράσεις).
 - Με χιλιάδες ιδιότητες: προβλήματα **ταχύτητας**, **ορθότητας** και **υπερεφαρμογής** με πολλούς αλγορίθμους μηχανικής μάθησης.
- Πόσο **αξίζει** κάθε υποψήφια ιδιότητα X ;
 - Αν ξέρουμε την τιμή της X , **πόση πληροφορία** κερδίζουμε;
 - Πόσο μειώνεται η **αβεβαιότητά** μας για τη σωστή απάντηση;
 - **$C = 1$** (ανεπιθύμητο) ή **$C = 0$** (επιθυμητό).
 - Πόσο μειώνεται η **αβεβαιότητά** μας για την τιμή της **τυχαίας μεταβλητής** C , αν ξέρουμε την τιμή της X ;

Παράδειγμα κωδικοποίησης τιμών

- Έστω οι ακόλουθες **δυνατές τιμές** της **C**:
 - c_1 με $P(c_1) = 1/4$, c_2 με $P(c_2) = 1/4$, c_3 με $P(c_3) = 1/2$.
- **Παράδειγμα** κωδικοποίησης:
 - Χρησιμοποιούμε **λιγότερα δυφία (bits)** για τις **πιο πιθανές** τιμές.
 - $c_1 \rightarrow 10$, $c_2 \rightarrow 11$. Χρησιμοποιούμε $-\log_2(1/4) = 2$ δυφία.
 - $c_3 \rightarrow 0$. Χρησιμοποιούμε $-\log_2(1/2) = 1$ δυφίο.
 - Αν ο παραλήπτης λάβει «1», ξέρει ότι ακολουθεί και δεύτερο δυφίο.
 - Αναμενόμενος αριθμός δυφίων: $1/4 \cdot 2 + 1/4 \cdot 2 + 1/2 \cdot 1 = 1.5$
- Γενικότερα, στην **ιδανική κωδικοποίηση** (ελάχιστος αναμενόμενος αριθμός μεταδιδόμενων δυφίων) χρησιμοποιούμε $-\log_2(P(c_i))$ **δυφία για κάθε τιμή c_i** .
 - Αν οι πιθανότητες $P(c_i)$ **δεν είναι δυνάμεις του 2**, ίσως είναι **αδύνατον** να χρησιμοποιήσουμε ακριβώς $-\log_2(P(c_i))$ **δυφία** για κάποια c_i στην πράξη. Αλλά δεν μας πειράζει αυτό, η ιδανική κωδικοποίηση είναι θεωρητική.
 - Στην **πράξη**, ο αναμενόμενος αριθμός μεταδιδόμενων δυφίων θα είναι γενικά **μεγαλύτερος ή ίσος** αυτού της ιδανικής κωδικοποίησης, τόσο μεγαλύτερος όσο απέχει η κωδικοποίηση από την ιδανική.

Εντροπία

Εδώ η **ορθή** απόκριση.

- **Εντροπία** της τυχαίας μεταβλητής C .
 - Δείχνει **πόσο αβέβαιοι** είμαστε για την **τιμή της C** .
 - Πόση είναι η **ελάχιστη ποσότητα πληροφορίας** που πρέπει να μας δοθεί για να γνωρίζουμε με βεβαιότητα την τιμή της C .
 - Ποιος είναι στην **καλύτερη περίπτωση** (με την ιδανική κωδικοποίηση) ο **αναμενόμενος αριθμός δυφίων** που πρέπει να μας μεταδοθεί για να καθοριστεί η τιμή της C ;

αναμενόμενη
τιμή

$$H(C) = - \sum_c P(C = c) \log_2 P(C = c)$$

Δυφία που
χρησιμοποιούμε
στην ιδανική
κωδικοποίηση για
την παράσταση
κάθε τιμής.

- Στην περίπτωση δύο κατηγοριών:

$$H(C) = -P(C = 1) \cdot \log_2 P(C = 1) - P(C = 0) \cdot \log_2 P(C = 0)$$

Εκτίμηση πιθανοτήτων από τα παραδείγματα εκπαίδευσης.

Παράδειγμα

- Συλλογή **800 μηνυμάτων** e-mail εκπαίδευσης.
 - Προηγούμενα μηνύματα που έχω λάβει.
 - **200 ανεπιθύμητα** διαφημιστικά μηνύματα (spam).
 - **600 επιθυμητά** μηνύματα.
- Εκτιμήστε την **εντροπία της C** (κατηγορία νέου μηνύματος) με βάση τα μηνύματα εκπαίδευσης.
 - **$C = 1$** (ανεπιθύμητο) ή **$C = 0$** (επιθυμητό).
 - $\log_2 3 = 1.585$
- **Πόση γίνεται η εντροπία της C** αν **όλα** τα μηνύματα εκπαίδευσης είναι **επιθυμητά** (ή ανεπιθύμητα);
- Αν είναι **400 επιθυμητά** και **400 ανεπιθύμητα**;

$$H(C) = -P(C = 1) \cdot \log_2 P(C = 1) - P(C = 0) \cdot \log_2 P(C = 0)$$

Κέρδος πληροφορίας

- Εντροπία της C αν μάθουμε ότι η τιμή της X είναι 1:

$$H(C | X = 1) = - \sum_c P(C = c | X = 1) \cdot \log_2 P(C = c | X = 1)$$

- Εντροπία της C αν μάθουμε ότι η τιμή της X είναι 0:

$$H(C | X = 0) = - \sum_c P(C = c | X = 0) \cdot \log_2 P(C = c | X = 0)$$

- **Κέρδος πληροφορίας IG** (information gain):

Αναμενόμενη μείωση της $H(C)$, αν μάθουμε την τιμή της X .

$$IG(C, X) = IG(X, C) = H(C) - \sum_x P(X = x) \cdot H(C | X = x)$$

αναμενόμενη τιμή \rightarrow x για κάθε δυνατή τιμή του X \rightarrow $H(C | X = x)$

Επιλογή ιδιοτήτων με IG

- Υπολογίζουμε το κέρδος πληροφορίας $IG(C, X)$ που παρέχει κάθε υποψήφια ιδιότητα X .
 - Π.χ. η X μπορεί να παριστάνει την ύπαρξη ($X = 1$) ή μη ($X = 0$) της λέξης «money» στο μήνυμα.
 - Οι πιθανότητες εκτιμούνται χρησιμοποιώντας τα παραδείγματα εκπαίδευσης.
- Επιλέγουμε τις ιδιότητες με τα m υψηλότερα $IG(C, X)$.
 - Π.χ. $m = 500$. Συχνά αφού παραλείψουμε πολύ συχνές λέξεις (stop-words, π.χ. άρθρα, συνδέσμους) και πολύ σπάνιες λέξεις.
- Παριστάνουμε τα παραδείγματα εκπαίδευσης και τα νεοεισερχόμενα μηνύματα ως διανύσματα m διαστάσεων.
 - Π.χ. $\langle X_1, X_2, X_3, \dots, X_m \rangle = \langle 0, 1, 1, \dots, 0 \rangle$

Παράδειγμα επιλεγμένων ιδιοτήτων

Λέξη του X_i	$P(X_i=1)$	$P(X_i=1 C=0)$	$P(X_i=1 C=1)$
!	0.484105	0.216129	0.828157
\$	0.257947	0.040322	0.538302
language	0.247956	0.440322	0.002070
money	0.163487	0.001612	0.372670
remove	0.146230	0.001612	0.333333
free	0.309718	0.104838	0.573498
university	0.219800	0.374193	0.022774

Προεπεξεργασία κειμένων

our highly successful multi – national company gives you an exclusive business that generates an extra weekly income of up to \$ 600 or more ... anyone can easily make money ... if you wish to be removed from our list ...

< money:1, language:0, natural:0, \$:1, adult:0, call:0, exclusive:1, **successful:1**, removed:1, **generates:1**, ...>

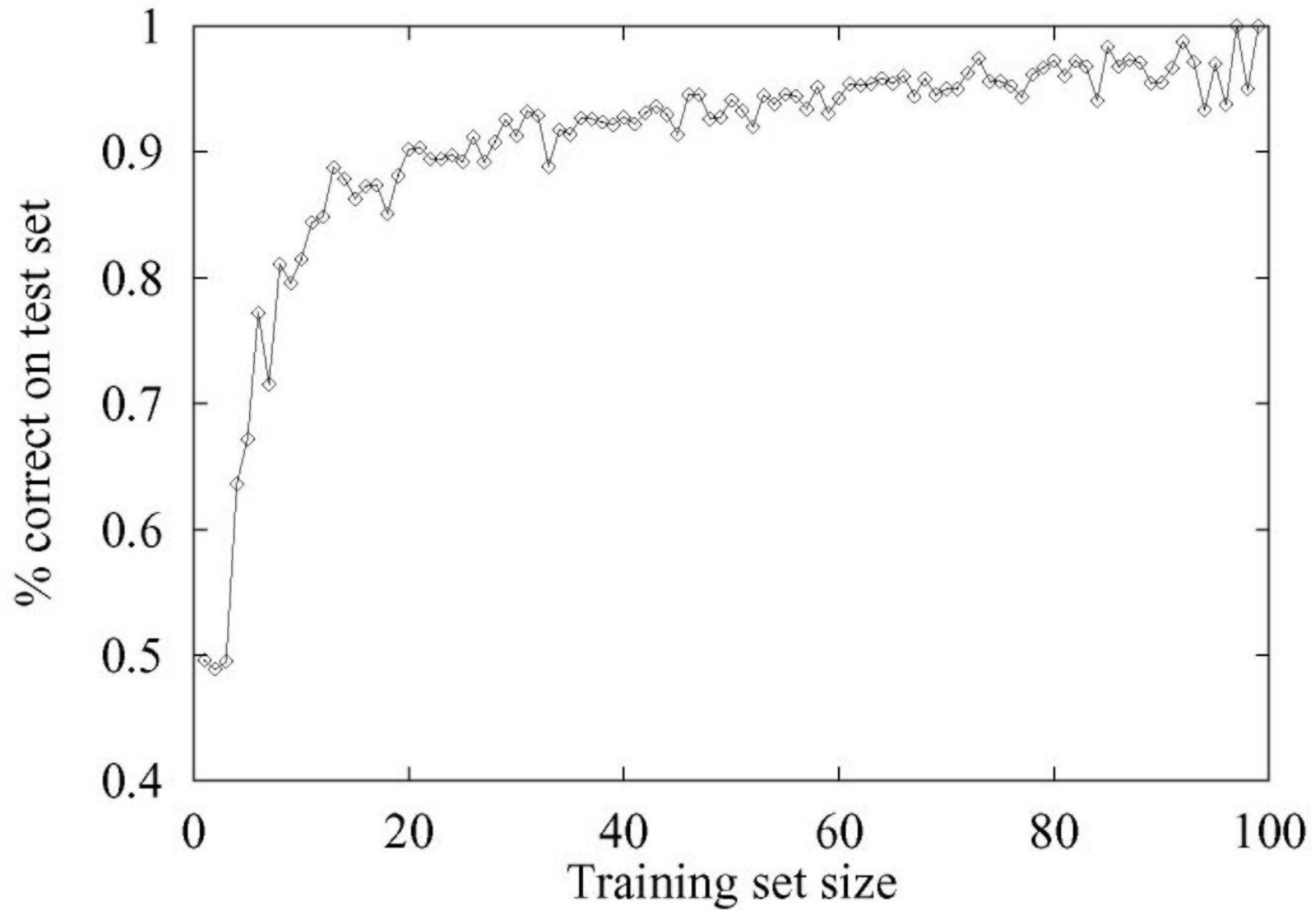
call for papers 9 th european workshop on natural language generation ... is a subfield of natural language processing that generates texts in human languages from non-linguistic data or knowledge ... for the systems to be successful ...

< money:0, language:1, natural:1, \$:0, adult:0, call:1, exclusive:0, **successful:1**, removed:0, **generates:1**, ...>

Δεδομένα εκπαίδευσης, ανάπτυξης, ελέγχου

- **Δεδομένα εκπαίδευσης** (training data):
 - Δεδομένα στα οποία εκπαιδεύεται ο αλγόριθμος μάθησης.
 - Συχνά δοκιμάζουμε να εκπαιδεύσουμε τον αλγόριθμο στο $x\%$ των δεδομένων εκπαίδευσης ($x = 10\%, \dots, 100\%$).
- **Δεδομένα ανάπτυξης/ελέγχου** (development/test data):
 - Δεδομένα στα οποία ελέγχουμε την επίδοση της συνάρτησης που μάθαμε, κατά την ανάπτυξη/τελική δοκιμή.
 - **Διαφορετικά** από τα δεδομένα εκπαίδευσης, αλλά από τον ίδιο πληθυσμό (π.χ. κριτικές ίδιων ειδών προϊόντων).
- **Παραδείγματα μέτρων αξιολόγησης:**
 - **Ποσοστό ορθότητας** (accuracy): ποσοστό περιπτώσεων για τις οποίες η απόκριση του ταξινομητή είναι σωστή.
 - **Ποσοστό λάθους** (error rate): $1 - \text{accuracy}$.
 - **Μέσο τετραγωνικό ή απόλυτο σφάλμα**, όταν προβλέπουμε τιμή (παλινδρόμηση) αντί για κατηγορία.

Καμπύλη μάθησης



Αναζήτηση υπο-συνόλων ιδιοτήτων

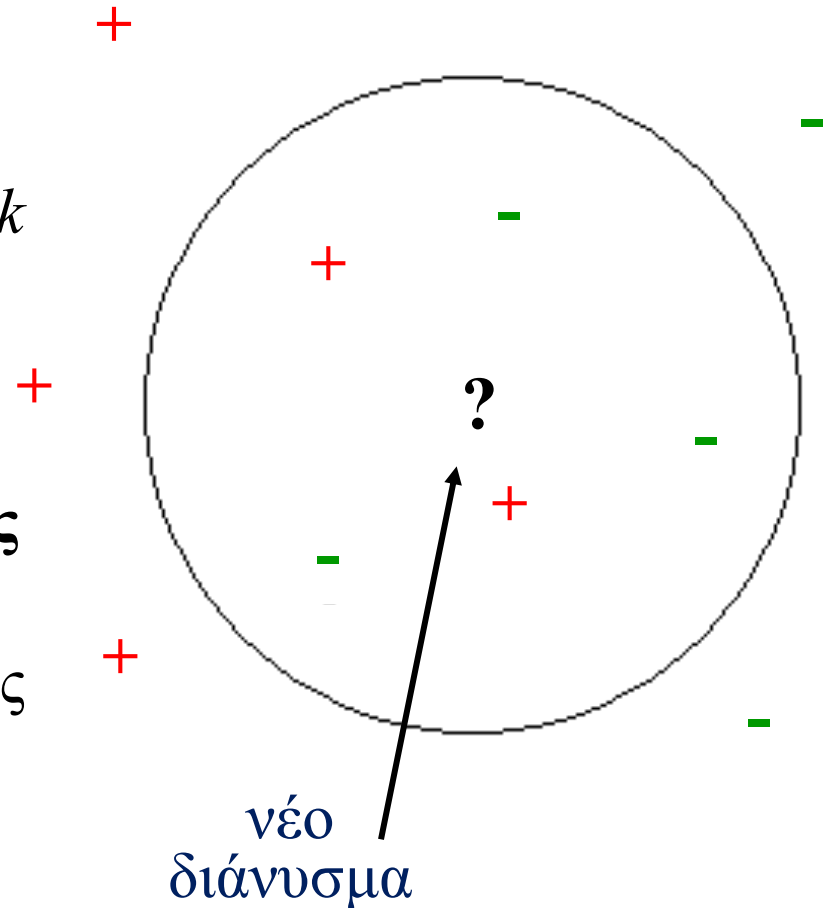
- Ως τώρα θεωρούσαμε ότι **κάθε υποψήφια ιδιότητα αξιολογείται ξεχωριστά.**
 - Με *IG*, υπάρχουν και άλλα μέτρα (π.χ. χ^2).
- Μπορεί, όμως, μια ιδιότητα ουσιαστικά να παρέχει περίπου τις **ίδιες πληροφορίες με μια άλλη.**
 - Π.χ. αν οι λέξεις «money» και «rich» εμφανίζονται (ή όχι) σχεδόν πάντα μαζί.
 - Οπότε μπορεί και οι δύο να έχουν υψηλό **IG**, αλλά η μία να **περιττεύει.**
- Μπορεί, επίσης, δύο (ή περισσότερες) ιδιότητες **από μόνες τους να μην είναι πολύ χρήσιμες**, αλλά και οι δύο (ή περισσότερες) **μαζί να προβλέπουν καλά** την κατηγορία.
 - Π.χ. υψηλός πυρετός ΚΑΙ δύσπνοια...

Αναζήτηση υπο-συνόλων ιδιοτήτων – συνέχεια

- Μπορούμε να θεωρήσουμε το πρόβλημα **επιλογής του καλύτερου υποσυνόλου ιδιοτήτων** (μεταξύ των διαθέσιμων) ως ένα **πρόβλημα αναζήτησης**.
 - Π.χ. **αρχική κατάσταση**: όλες οι υποψήφιες ιδιότητες.
 - **Τελεστής μετάβασης**: αφαίρεση μιας υποψήφιας ιδιότητας.
 - **Συνάρτηση αξιολόγησης κατάστασης**: ποσοστό ορθότητας που επιτυγχάνουμε με το υπο-σύνολο ιδιοτήτων της κατάστασης εκπαιδευοντας στα δεδομένα εκπαίδευσης και αξιολογώντας σε διαφορετικά δεδομένα ανάπτυξης (ή «επικύρωσης»).
 - Αναζήτηση π.χ. με **αναρρίχηση λόφου, beam search...**
- Εναλλακτικά μπορούμε να χρησιμοποιήσουμε **μείωση διαστάσεων** μέσω **Singular Value Decomposition (SVD)**.
 - Βλ. μάθημα «Υπολογιστικά Μαθηματικά».
 - Απεικονίζει τα διανύσματα σε νέο χώρο λιγότερων διαστάσεων.
 - **Δεν επιλέγουμε** απλά ιδιότητες, **κατασκευάζουμε νέες** (λιγότερες) που **συνδυάζουν πληροφορίες** των αρχικών.

Αλγόριθμος k κοντινότερων γειτόνων (k -NN)

- Κατά την **εκπαίδευση** απλά **αποθηκεύουμε** τα διανύσματα εκπαίδευσης.
- Κατά τη **χρήση**, βρίσκουμε στα παραδείγματα εκπαίδευσης τους k κοντινότερους **γείτονες** του νέου διανύσματος (π.χ. $k = 5$).
- Κατατάσσουμε το νέο διάνυσμα στην κατηγορία της **πλειοψηφίας** των γειτόνων (εδώ «-»).
- Σε προβλήματα που οι αποκρίσεις είναι **πραγματικοί** αριθμοί, επιστρέφουμε π.χ. τη **μέση τιμή** των γειτόνων.



Μέτρο απόστασης

- Παραδείγματα μέτρων απόστασης:
 - Ευκλείδεια απόσταση
 - ή για δίτιμες ιδιότητες, σε πόσες θέσεις διαφέρουν τα δύο διανύσματα.

$$d(\vec{x}_i, \vec{x}_j) \equiv \sum_{r=1}^m \delta(x_{ir}, x_{jr})$$

$$\delta(x, y) \equiv \begin{cases} 0, & \text{αν } x = y \\ 1, & \text{διαφορετικά} \end{cases}$$

- Βελτίωση: οι διαφορές **σταθμίζονται** ανάλογα με το πόσο σημαντική είναι η ιδιότητα (attribute weighting):

$$d(\vec{x}_i, \vec{x}_j) \equiv \sum_{r=1}^m IG(C, X_r) \cdot \delta(x_{ir}, x_{jr})$$

Παράδειγμα

- Έχουν αποθηκευτεί τα ακόλουθα διανύσματα εκπαίδευσης:

$\langle 1, 0, 0, 1, \mathbf{1} \rangle$

$\langle 1, 1, 0, 1, \mathbf{1} \rangle$

$\langle 0, 0, 1, 1, \mathbf{0} \rangle$

$\langle 0, 0, 1, 0, \mathbf{0} \rangle$

Ο τελευταίος αριθμός είναι η κατηγορία.

- Φτάνει **νέο** διάνυσμα:

$\langle 1, 1, 0, 0, ? \rangle$

- Σε **ποια κατηγορία** κατατάσσεται το νέο διάνυσμα;
 - Για $k = 3$ και με ισοβαρείς ιδιότητες.

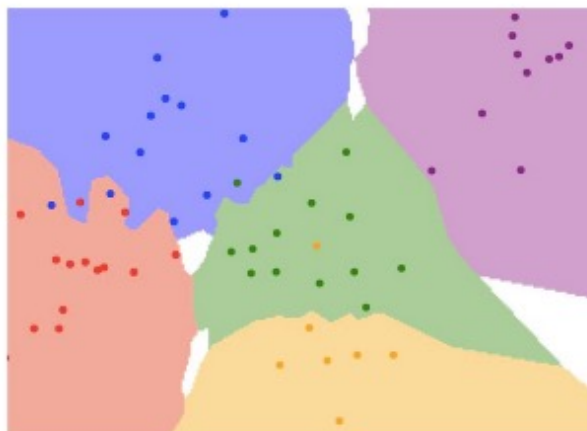
Η επίδραση του k

K-Nearest Neighbors

Instead of copying label from nearest neighbor, take **majority vote** from K closest points



$K = 1$



$K = 3$



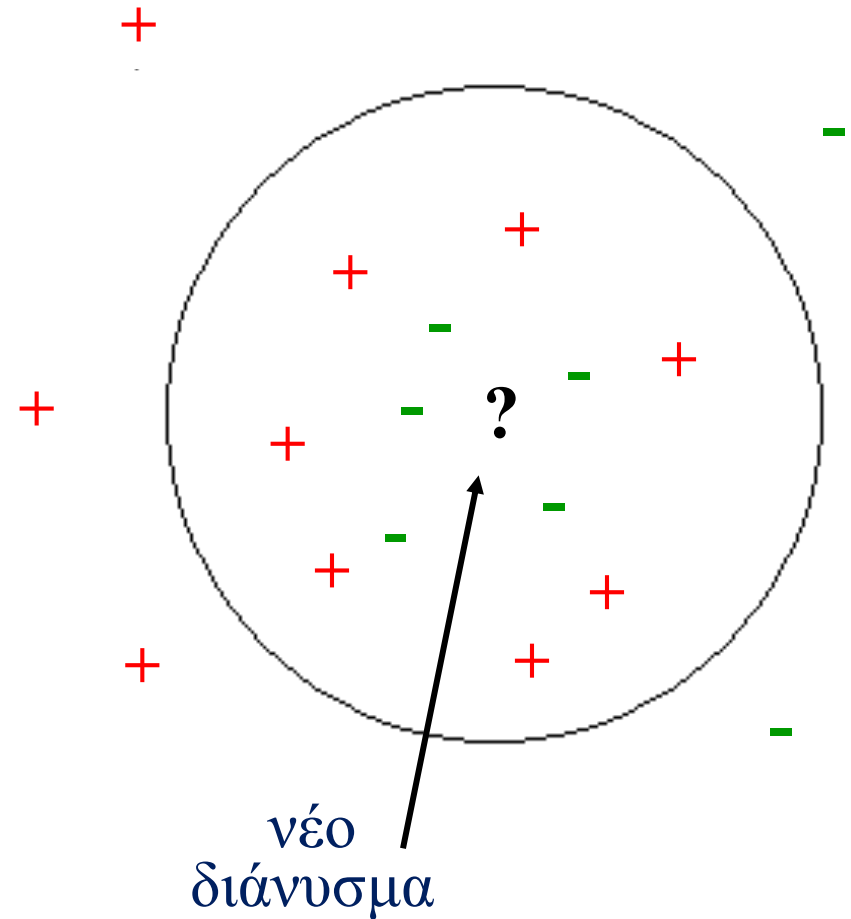
$K = 5$

Βάρος απόστασης (distance weighting)

Περαιτέρω βελτίωση:

- Η ψήφος κάθε γείτονα \vec{x}_i έχει **βάρος** π.χ. αντιστρόφως ανάλογο της **απόστασής** του από το νέο διάνυσμα \vec{x} .
- Δηλαδή αφού βρούμε τους k κοντινότερους γείτονες, πολλαπλασιάζουμε την ψήφο κάθε γείτονα \vec{x}_i με:

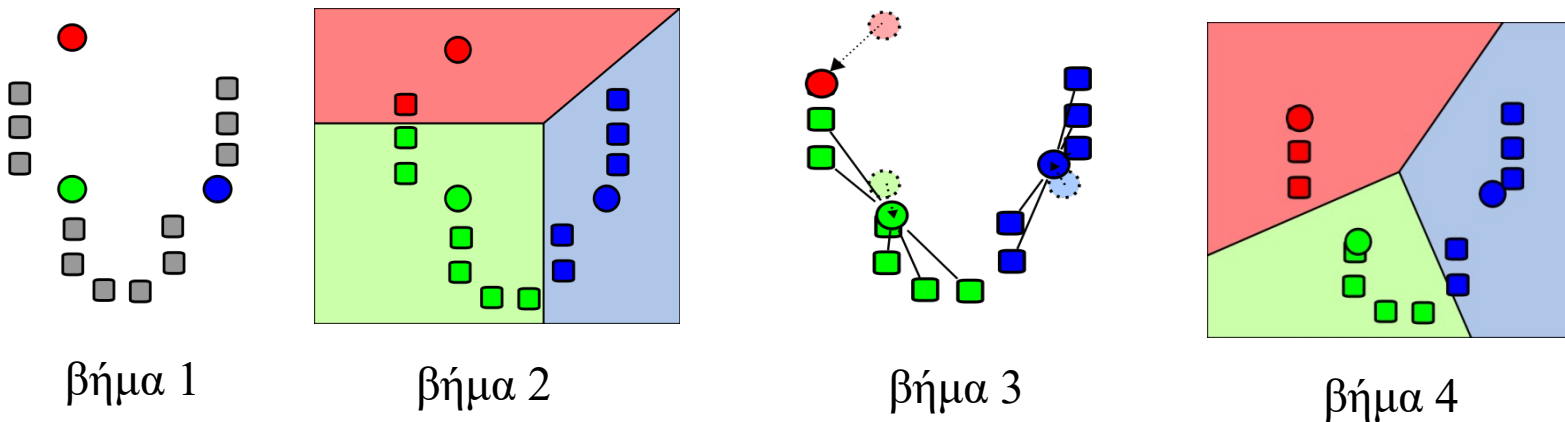
$$1/d(\vec{x}_i, \vec{x})$$



Χαρακτηριστικά k -NN

- Πολύ μικρό υπολογιστικό κόστος εκπαίδευσης.
 - Απλά αποθηκεύουμε τα παραδείγματα εκπαίδευσης.
- Αλλά μεγάλες απαιτήσεις μνήμης.
 - Αποθηκεύουμε όλα τα παραδείγματα εκπαίδευσης.
- Και μεγάλο υπολογιστικό κόστος κατά τον υπολογισμό της αποκρίσεως.
 - Στην απλή μορφή, χρειάζεται να υπολογίζουμε κάθε φορά τις αποστάσεις από όλα τα παραδείγματα εκπαίδευσης.
 - Υπάρχουν προσεγγίσεις που μειώνουν πολύ αυτό το κόστος.
- Μπορεί να προσεγγίσει οποιαδήποτε συνάρτηση.
- Δεν μπορούμε να παραστήσουμε άμεσα σε λογική τη γνώση που αποκτά.

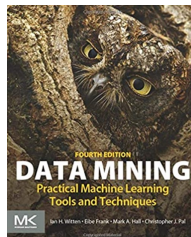
Ομαδοποίηση με τον k -means



- Ξεκίνα με k τυχαία κέντρα βαρών (ένα για κάθε ομάδα).
- Θεώρησε ότι κάθε στιγμιότυπο ανήκει στην ομάδα με το κοντινότερο κέντρο βάρους.
- Υπολόγισε το (νέο) κέντρο βάρους κάθε ομάδας.
- Επαναλήψεις ως σύγκλιση...
- Στην πράξη χρησιμοποιείται πολύ. Πρέπει να ξέρουμε το k .
- Προσπαθεί να ελαχιστοποιήσει τις **συνολικές αποστάσεις** των στιγμιότυπων **από τα κέντρα βαρών** των ομάδων τους.
- Μπορεί να παγιδευτεί σε τοπικό ελάχιστο. Ευαίσθητος στα αρχικά κέντρα βαρών. **Επανεκκινήσεις** με διαφορετικά αρχικά κέντρα βαρών.

Βιβλιογραφία

- Russel & Norvig (4^η έκδοση): ενότητες 19.3.3, 19.7.1.
 - Θα καλύψουμε αρκετές άλλες ενότητες του κεφαλαίου 19 στις επόμενες διαλέξεις.
- Βλαχάβας κ.ά: υπο-ενότητα «Κριτήριο διαχωρισμού με βάση την εντροπία» της ενότητας 18.5.1, ενότητα 18.7, υπο-ενότητα «Αλγόριθμος k-μέσων» της ενότητας 18.13.2.
- Υλοποιήσεις σε Java των περισσότερων αλγορίθμων μάθησης που θα συναντήσουμε περιλαμβάνονται στο σύστημα Weka.
 - Διατίθεται ελεύθερα (<http://www.cs.waikato.ac.nz/ml/weka/>).
 - Το βιβλίο των I.H. Witten και E. Frank «Data Mining: Practical Machine Learning Tools and Techniques», 4^η έκδοση, 2016, Morgan Kaufmann είναι μια καλή εισαγωγή στη μηχανική μάθηση (και την εξόρυξη γνώσεων). Περιλαμβάνει κεφάλαια για τη χρήση του Weka. Υπάρχει στη βιβλιοθήκη.
- Στα φροντιστήρια θα μάθετε και για το scikit-learn (Python).
 - <https://scikit-learn.org/>



Βιβλιογραφία – συνέχεια

- Περισσότερα για το PCA (και SVD) θα βρείτε στην ενότητα 12.2 του βιβλίου «Machine Learning – A Probabilistic Perspective», του K.P. Murphy, MIT Press, 2012. (Υπάρχει στη βιβλιοθήκη του ΟΠΑ.)
- Δείτε προαιρετικά και το εισαγωγικό βιβλίο «A Course in Machine Learning» του Hal Daumé III.
 - Διατίθεται ελεύθερα. Βλ. <http://ciml.info/>
- Για μεθόδους πολύ γρήγορης προσεγγιστικής ανάκτησης κοντινότερων γειτόνων, δείτε π.χ. τα:
 - <http://erikbern.com/2015/10/01/nearest-neighbors-and-vector-models-part-2-how-to-search-in-high-dimensional-spaces/>
 - <https://github.com/spotify/annoy>

