Εργαστήριο Συστημάτων Τεχνητής Νοημοσύνης και Μάθησης
Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Ανασκόπηση στοιχείων

Τζούβελη Παρασκευή

Γραμμική Άλγεβρα

Τανυστές

- **Τανυστές μηδενικής τάξης - Βαθμωτά (Scalars)**
  - ένας μόνο αριθμός ($s \in \Re$)
- **Τανυστές πρώτης τάξης - Διανύσματα (Vectors)**
  - μονοδιάστατος πίνακας όπου $x_1 .. x_n \in \Re$

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ x_n \end{bmatrix}$$

- **Τανυστές δεύτερης τάξης - Πίνακες (Matrices)**
  - δισδιάστατοι πίνακες, με $a_{11} ... a_{nn} \in \Re$ και

$$A \in \Re^{mxn}, A = \begin{bmatrix} a_{11} & a_{12} & a_{1n} \\ . & . & . \\ a_{m1} & a_{m2} & a_{mn} \end{bmatrix}$$

  -
- **Τανυστές μεγαλύτερης από 2ης τάξης - Τανυστές (Tensors)**
  - πολυδιάστατοι πίνακες π.χ. 3ης τάξης όπου $b_{111} ... b_{kmn} \in \Re$ και $B \in \Re^{kxmxn}$

Ανάστροφο διάνυσμα x^T και πίνακας A^T

- Ορίζουμε τον **ανάστροφο (transpose) του διανύσματος x** και συμβολίζουμε **x^T**, ως τον πίνακα γραμμή N θέσεων με στοιχεία: $x^T = [\, x_1 \quad x_2 \quad .. \quad x_n \,]$

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ x_n \end{bmatrix}$$

$$(\mathbf{A}^T)_{i,j} = (\mathbf{A})_{j,i}$$

- Ορίζουμε τον **ανάστροφος πίνακα** του πίνακα $A = (a_{ij}) \in M_{mxn}$ και συμβολίζουμε **A^T**, ως τον πίνακα που προκύπτει από τον Α με εναλλαγή μεταξύ γραμμών και στηλών του:

$$\mathbf{A} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \\ A_{31} & A_{32} \end{bmatrix} \quad \mathbf{A}^T = \begin{bmatrix} A_{11} & A_{21} & A_{31} \\ A_{12} & A_{22} & A_{32} \end{bmatrix}$$

Κύρια διαγώνιος

Ανάστροφος πίνακας A^T

Παραδείγματα

$$A=\begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix}_{mxn}$$

$$C = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 5 & -6 \\ -4 & 3 & 0 \\ -1 & 8 & 9 \end{bmatrix}_{4x3}$$

$$A^T = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,m} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,m} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,m} \end{bmatrix}_{nxm}$$

$$C^T = \begin{bmatrix} 1 & 2 & -4 & -1 \\ 0 & 5 & 3 & 8 \\ -1 & -6 & 0 & 9 \end{bmatrix}_{3x4}$$

Ιδιότητες

$$(A^T)^T = A$$

$$(A+B)^T = A^T + B^T$$

$$(AB)^T = B^T A^T$$

Τετραγωνικοί Πίνακες

Ένας πίνακας **A** είναι **<u>τετραγωνικού τύπου n</u>** όταν έχει διάσταση **nxn** π.χ. $A=\begin{pmatrix} 5 & 2 \\ 4 & 6 \end{pmatrix}$ είναι τετραγωνικού τύπου 2.

Ένας **τετραγωνικός πίνακας** με γραμμικά **εξαρτημένες** στήλες (ή γραμμές) είναι γνωστός ως **μη αντιστρέψιμος** ή **ιδιόμορφος (singular)**.

→ Αν οι στήλες (ή οι γραμμές) είναι **γραμμικά εξαρτημένες**, *τότε ο πίνακας δεν έχει αντίστροφο και η ορίζουσα του είναι 0.*

Ένας **τετραγωνικός πίνακας** με **γραμμικά ανεξάρτητες** στήλες (ή γραμμές) ονομάζεται **αντιστρέψιμος** (**non-singular**).

→ Σε αυτή την περίπτωση, *η ορίζουσα του πίνακα είναι διαφορετική από το μηδέν.*

Διαγώνιοι Πίνακες

Ένας πίνακας $\mathbf{A} \in \mathbf{R}^{nxn}$ ονομάζεται **<u>διαγώνιος πίνακας</u>**, αν και μόνο αν $a_{ij}=0$ όταν $i \neq j$ δηλαδή όλα τα στοιχεία <u>έξω από την κύρια διαγώνιο</u> είναι 0

π.χ. $A = \begin{pmatrix} 5 & 0 \\ 0 & -3 \end{pmatrix}$

Μοναδιαίος Πίνακας Τύπου n, I_n

Ένας διαγώνιος πίνακας **$I \in R^{nxn}$** ονομάζεται **<u>μοναδιαίος πίνακας</u>** αν και μόνο αν

$$(I_n)_{ij} = \delta_{ij} = \begin{cases} 1, \text{ για } i=j \\ 0, \text{ για } i \neq j \end{cases}$$

Ίχνος Τετραγωνικού Πίνακας

Έστω $A \in$ **R^{nxn}** , τότε το **<u>ίχνος του A</u>** ορίζεται ως : $tr(A) = trace(A) : \sum_{i=1}^{n} a_{ii} = a_{11} + a_{22} + \ldots a_{nn}$

δηλαδή ίσο με το άθροισμα των στοιχείων της κύριας διαγωνίου.

Κάτω Τριγωνικός Πίνακας

Ένας πίνακας **A**$\in$ **R**^{nxn} ονομάζεται **<u>κάτω τριγωνικός</u>** αν και μόνο αν a_{ij} =0 όταν i<j , δηλαδή όλα τα στοιχεία πάνω από την κύρια διαγώνιο είναι 0

$$\begin{pmatrix} 3 & 0 & 0 \\ 4 & 6 & 0 \\ 5 & 7 & 8 \end{pmatrix}$$

Άνω Τριγωνικός Πίνακας

Ένας πίνακας **A**$\in$ **R**^{nxn} ονομάζεται **<u>άνω τριγωνικός</u>** αν και μόνο αν a_{ij} =0 όταν i>j , δηλαδή όλα τα στοιχεία κάτω από την κύρια διαγώνιο είναι 0

$$\begin{pmatrix} 3 & 4 & 5 \\ 0 & 6 & 7 \\ 0 & 0 & 8 \end{pmatrix}$$

Αντίστροφος πίνακας

Ο πίνακας **A∈ R^{nxn}** είναι **<u>αντιστρέψιμος</u>** εάν υπάρχει ένας πίνακας **B ∈ R^{nxn}** έτσι ώστε **B·A = I** και **A·B = I**, όπου **I** είναι nxn μοναδιαίος πίνακας.

Υπάρχει το πολύ ένας τέτοιος **B** και λέγεται **<u>αντίστροφος</u>** του **A** και συμβολίζεται με **A^{-1}** : **A·A^{-1} = A^{-1} ·A** = I

→ Οπότε, για να είναι ο πίνακας **A∈ R^{nxn}** αντιστρέψιμος θα πρέπει **η ορίζουσά του είναι διάφορη του 0** : **detA ≠ 0**

Ιδιότητες

$$AA^{-1} = I = A^{-1}A$$

$$(AB)^{-1} = B^{-1}A^{-1}$$

$$(A+B)^{-1} \neq A^{-1} + B^{-1}$$

Αντίστροφος πίνακας

- **Παράδειγμα 1.1 (Υπαρξη αντιστρόφου πίνακα 2x2)** Θεωρούμε έναν πίνακα $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \in \mathbb{R}^{2x2}$. Αν τον πολλαπλασιάσουμε με τον πίνακα $B = \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \in \mathbb{R}^{2x2}$
- τότε θα έχουμε $= (a_{11}a_{22} - a_{12}a_{21})\mathbf{I}$
  Τότε ο αντίστροφος $^{-1}$ του πίνακα θα είναι ο

$$A^{-1} = \frac{1}{(a_{11}a_{22} - a_{12}a_{21})} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \tag{1.1}$$

αν και μόνο αν $(a_{11}a_{22} - a_{12}a_{21}) \neq 0$ (όπου $(a_{11}a_{22} - a_{12}a_{21}$ είναι η ορίζουσα του πίνακα A. Μπορούμε να χρησιμοποιούμε την ορίζουσα ενός πίνακα για να αποφανθούμε αν αυτός είναι αντιστρέψιμος.

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad ad - bc \neq 0 \qquad \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Συμμετρικοί πίνακες

- Ένας **τετραγωνικός** πίνακας S ∈ R^{nxn} ο οποίος είναι **ίσος** με τον **ανάστροφό** του, δηλαδή $S = S^T$, είναι ένας **<u>συμμετρικός πίνακας</u>**.

  π.χ. εάν A είναι ένας **πίνακας μέτρησης αποστάσεων**, με $A_{i,\,j}$ που δίνει την απόσταση από το σημείο i στο σημείο j, τότε $A_{i,\,j} = A_{j,\,i}$ επειδή οι συναρτήσεις απόστασης είναι συμμετρικές.

$$\begin{pmatrix} 3 & 5 \\ 5 & -1 \end{pmatrix} \qquad \begin{pmatrix} 7 & -5 & 2 \\ -5 & 3 & 4 \\ 2 & 4 & 8 \end{pmatrix} \qquad \begin{pmatrix} 5 & 6 & -4 & -3 \\ 6 & 7 & 2 & 9 \\ -4 & 2 & 1 & 5 \\ -3 & 9 & 5 & 6 \end{pmatrix}$$

Αντισυμμετρικοί πίνακες

- Αν αντίθετα, ο **τετραγωνικός** πίνακας $\mathbf{S} \in \mathbf{R}^{nxn}$ ο οποίος είναι **ίσος** με τον αρνητικό του ανάστροφο, δηλαδή $\mathbf{S = S^{-T}}$ τότε λέμε ότι ο S είναι ένας **αντισυμμετρικός πίνακας**.

Δηλαδή: $\mathbf{(A)^{ij}=-(A)^{ji}, \forall\ i,j}$

Παραδείγματα

$$\begin{pmatrix} 0 & 3 \\ -3 & 0 \end{pmatrix} \qquad \begin{pmatrix} 0 & 5 & 2 \\ -5 & 0 & 4 \\ -2 & -4 & 0 \end{pmatrix} \qquad \begin{pmatrix} 0 & 6 & -4 & 3 \\ -6 & 0 & 2 & 9 \\ 4 & -2 & 0 & -5 \\ -3 & -9 & 5 & 0 \end{pmatrix}$$

Πράξεις πινάκων

- **Πρόσθεση πινάκων**

  Αν οι πίνακες έχουν το ίδιο μέγεθος μπορούμε να προσθέτουμε τα αντίστοιχα στοιχεία τους:

  $C{=}A{+}B$ όπου $C_{i,j} = A_{i,j} + B_{i,j}$

- **Πίνακας με βαθμωτό ή διάνυσμα**

  Προσθέτουμε/πολλαπλασιάζουμε το βαθμωτό με κάθε στοιχείο του πίνακα:

  D = α · B + c όπου $D_{i,j} = a \cdot B_{i,j} + c$

- **Πολλαπλασιασμός πινάκων**

  Προϋπόθεση: Οι στήλες του A ίσες με τις γραμμές του B

  $$C = A * B, C_{i,j} = \sum_k (A_{i,j} \cdot B_{i,j})$$

  $$C(mxp) = A(mxn) * B(nxp)$$

- **Γινόμενο στοιχείο προς στοιχείο (Element-wise product ή Hadamard product)**

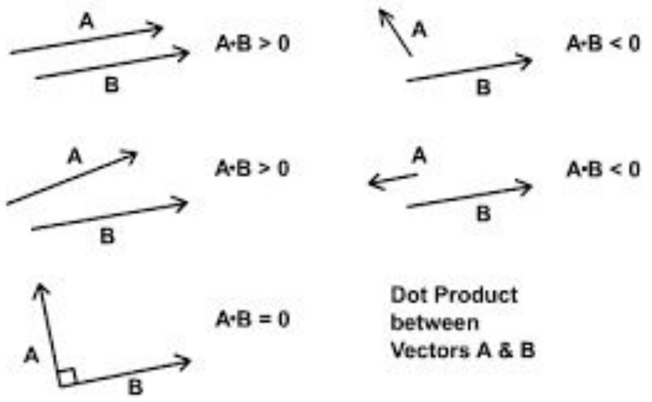
  $$A \cdot B = a_{i,j} \cdot b_{i,j}$$

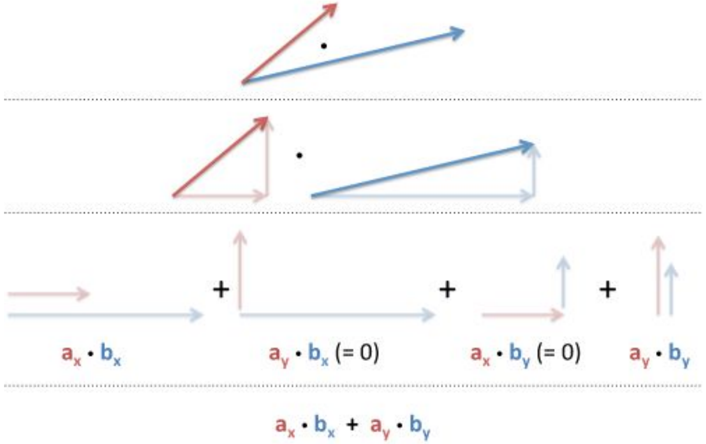
Εσωτερικό Γινόμενο Διανυσμάτων

Η ποσότητα $x^T y$ είναι το εσωτερικό γινόμενο των διανυσμάτων x και y του $\Re^n$

Αν τα x,y είναι ορθογώνια, $x^{\mathrm{T}} y = 0$

$$\begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} * \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix} = x_1 * y_1 + x_2 * y_2 \dots + x_n * y_n$$





Γραμμική Ανεξαρτησία Διανυσμάτων

Εάν ο τετριμμένος συνδυασμός είναι ο μόνος που παράγει το μηδέν δηλαδή $c_1 u_1 + c_2 u2 + \ldots c_n u_n = 0$ συμβαίνει μόνον όταν $c_1 = c_2 = \ldots c_n = 0$, τότε τα διανύσματα $u_1, u_2, \ldots u_n$ είναι **<u>γραμμικώς ανεξάρτητα</u>**.

Πχ.

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \end{bmatrix} * \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 0$$

Αν αυτό δεν συμβαίνει, τότε είναι **<u>γραμμικώς εξαρτημένα</u>** και κάποιο από αυτά είναι γραμμικός συνδυασμός των υπολοίπων.

πχ.

$$\begin{aligned} c_1 - c_2 + 2c_3 &= 0 \\ c_2 + 3/2 c_3 &= 0 \end{aligned}$$

$$\mathrm{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 3 \\ 3 \end{bmatrix}, u_2 = \begin{bmatrix} -1 \\ 1 \\ -1 \\ -3 \end{bmatrix}, u_3 = \begin{bmatrix} 2 \\ 5 \\ 9 \\ 6 \end{bmatrix} \rightarrow c_1 \begin{bmatrix} 1 \\ 1 \\ 3 \\ 3 \end{bmatrix} + c_2 \begin{bmatrix} -1 \\ 1 \\ -1 \\ -3 \end{bmatrix} + c_3 \begin{bmatrix} 2 \\ 5 \\ 9 \\ 6 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 & -1 & 2 \\ 0 & 1 & 3/2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Γραμμική Ανεξαρτησία Διανυσμάτων

Άσκηση Εξετάστε αν τα διανύσματα $(1,1,0,0)$, $(1,0,1,0)$, $(0,0,1,1)$, $(0,1,0,1)$ είναι γραμμικά ανεξάρτητα ή όχι και ελέγξτε εάν το διάνυσμα $(0,0,0,1)$ βρίσκεται στο χώρο που παράγουν.

$$A=\begin{bmatrix}1&1&0&0\\1&0&0&1\\0&1&1&0\\0&0&1&1\end{bmatrix}\longrightarrow\begin{bmatrix}1&1&0&0\\0&-1&0&1\\0&1&1&0\\0&0&1&1\end{bmatrix}\longrightarrow\begin{bmatrix}1&1&0&0\\0&-1&0&1\\0&0&1&1\\0&0&1&1\end{bmatrix}\longrightarrow$$

$$\longrightarrow\begin{bmatrix}1&1&0&0\\0&-1&0&1\\0&0&1&1\\0&0&0&0\end{bmatrix}=U$$

, άρα $r(A)=[$πλήθος μη-μηδεν. γραμμών του $U]$

$\Rightarrow r(A)=3$ → Πλήθος βασικών μεταβλητών

και $\dim N(A)=n-r(A)=4-3=1\neq 0$ → Πλήθος ελευθέρων μεταβλητών

άρα $\exists\ c\neq 0$ τ.ω. $Ac=0$ και τα διανύσματα είναι γρ. εξαρτημένα

Γραμμική Ανεξαρτησία Διανυσμάτων

Άσκηση Εξετάστε αν τα διανύσματα $(1,1,0,0)$, $(1,0,1,0)$, $(0,0,1,1)$, $(0,1,0,1)$ είναι γραμμικά ανεξάρτητα ή όχι και ελέγξτε εάν το διάνυσμα $(0,0,0,1)$ βρίσκεται στο χώρο που παράγουν.

Το διάνυσμα $(0,0,0,1)$ βρίσκεται στο χώρο που παράγουν ανν $\exists$ $\lambda_1, \lambda_2, \lambda_3, \lambda_4 \in \mathbb{R}$ τ.ω.

$$\begin{bmatrix}0\\0\\0\\1\end{bmatrix} = \lambda_1\begin{bmatrix}1\\1\\0\\0\end{bmatrix} + \lambda_2\begin{bmatrix}1\\0\\1\\0\end{bmatrix} + \lambda_3\begin{bmatrix}0\\0\\1\\1\end{bmatrix} + \lambda_4\begin{bmatrix}0\\1\\0\\1\end{bmatrix}$$

$$\Leftrightarrow \begin{bmatrix}1&1&0&0\\1&0&0&1\\0&1&1&0\\0&0&1&1\end{bmatrix}\begin{bmatrix}\lambda_1\\\lambda_2\\\lambda_3\\\lambda_4\end{bmatrix} = \begin{bmatrix}0\\0\\0\\1\end{bmatrix} \Leftrightarrow \begin{bmatrix}1&1&0&0\\0&-1&0&1\\0&0&1&1\\0&0&0&0\end{bmatrix}\begin{bmatrix}\lambda_1\\\lambda_2\\\lambda_3\\\lambda_4\end{bmatrix} = \begin{bmatrix}0\\0\\0\\1\end{bmatrix}$$

η τελευταία γραμμή δίνει $0\lambda_1 + 0\lambda_2 + 0\lambda_3 + 0\lambda_4 = 1$, άρα το σύστημα είναι ασύμβατο και επομένως,

$$(0,0,0,1) \notin \langle (1,1,0,0), (1,0,1,0), (0,0,1,1), (0,1,0,1) \rangle$$

Αν δεν υπάρχει λύση, τότε λέμε ότι το σύστημα είναι **ασύμβατο** (inconsistent).

Γραμμική Ανεξαρτησία Διανυσμάτων

Ένα σύνολο **n διανυσμάτων** του $\Re^m$ είναι κατ' ανάγκην εξαρτημένο, όταν **n>m**

Κάθε διάνυσμα **u** του διανυσματικού χώρου **V** μπορεί να εκφραστεί ως γραμμικός συνδυασμός του **w :** για κάποιους συντελεστές $\mathbf{c_i}$

$$u = c_1 w_1 + \ldots c_l w_l$$

Έστω διανύσματα που ξεκινούν από την αρχή ενός 3D χώρου:

2 εξαρτημένα διανύσματα ⇒ Περιέχονται στην ίδια ευθεία

3 εξαρτημένα διανύσματα ⇒ Περιέχονται στο ίδιο επίπεδο

Βάση και Διάσταση Διανυσματικού Χώρου

Βάση ενός διανυσματικού χώρου είναι ένα σύνολο διανυσμάτων που έχει ταυτόχρονα τις δύο ιδιότητες:

1. Είναι γραμμικώς ανεξάρτητο
2. Παράγει τον χώρο.

Όταν ένας τετραγωνικός πίνακας είναι αντιστρέψιμος, τότε οι στήλες του είναι ανεξάρτητες και αποτελούν μία βάση του .

Αν $u_1, .., u_m$ και $w_1, \ldots w_n$ είναι δύο βάσεις του ίδιου διανυσματικού χώρου **V,** τότε **m=n**

Το m (ή το n) εκφράζει τους "βαθμούς ελευθερίας" του χώρου και ονομάζεται **διάσταση** του V

→ **dim(V) = [μέγιστο πλήθος γραμμικώς ανεξάρτητων διανυσμάτων του V]**

π.χ. $\dim(R^2)=2$, $\dim(R^3)=3$

Βάση και Διάσταση Διανυσματικού Χώρου

Να βρεθεί μία βάση και η διάσταση του χώρου

$$V = \{(x, y, z) \in \Re^3) | x - y + 2z = 0\}$$

$$x - y + 2z = 0 \to x = y - 2z, y, z \in \Re$$

Οδηγοί

$$V = \{(x, y, z) \in \Re^3) | x = y - 2z\} = \{(y - 2z, y, z) | y, z \in \Re\}$$

$$= \{(y, y, 0) + (-2z, 0, z) | y, z \in \Re\}$$

$$= \{y(1, 1, 0) + z(-2, 0, 1) | y, z \in \Re\}$$

$$= < (1, 1, 0), (-2, 0, 1) >$$

e_1 , e_2

Για να αποτελέσουν βάση του χώρου αυτά τα 2 διανύσματα θα πρέπει να είναι γραμμικώς ανεξάρτητα:

$$\begin{bmatrix} 1 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix} \xrightarrow{1st*2+2nd} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix}$$

Ορθογώνιοι πίνακες

- Ένας τετραγωνικός πίνακας ονομάζεται **ορθογώνιος πίνακας** εάν και μόνο αν τα **n διανύσματα- στήλες** (ή n-γραμμές) του **αποτελούν ορθοκανονικό σύστημα του χώρου διάστασης nxn**

Π.χ. (ταυτοτικός) (μετάθεσης) (περιστροφή)

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \qquad \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

- Ισοδύναμα, ένας τετραγωνικός πίνακας **A** είναι **ορθογώνιος** αν **η μετάθεσή του (ανάστροφος) είναι ίση με τον αντίστροφό του** : $A^T = A^{-1}$

και ισχύει

$$A^T \mathrm{A} = A\mathrm{A}^T = I$$

Αντιστρέψιμος πίνακας: $(A^{-1})^T = (A^T)^{-1}$

Παράδειγμα

5) Αν $A \in \mathbb{R}^{n\times n}$ είναι αντιστρέψιμος, τότε: $\boxed{(A^{-1})^T = (A^T)^{-1}}$

π.χ. $A = \begin{bmatrix} 1 & -2 \\ -3 & 5 \end{bmatrix} \Rightarrow A^{-1} = \frac{1}{5-6}\begin{bmatrix} 5 & 2 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} -5 & -2 \\ -3 & -1 \end{bmatrix} \Rightarrow (A^{-1})^T = \begin{bmatrix} -5 & -3 \\ -2 & -1 \end{bmatrix}$

και $A^T = \begin{bmatrix} 1 & -3 \\ -2 & 5 \end{bmatrix} \Rightarrow (A^T)^{-1} = \frac{1}{-1}\begin{bmatrix} 5 & 3 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} -5 & -3 \\ -2 & -1 \end{bmatrix}$

Ορθογώνιοι πίνακες

- Η **ορίζουσα** οποιουδήποτε **ορθογώνιου πίνακα** είναι είτε +1 (κάνει μια καθαρή περιστροφή), ή **-1** (είναι μια καθαρή αντανάκλαση, ή μια σύνθεση της αντανάκλασης και της περιστροφής).

- Εάν $\mathbf{A} \in \mathbf{R}^{nxn}$ είναι ορθογώνιος πίνακας, τότε $\mathbf{A}^T \mathbf{A} = \mathbf{I}_n$

  Έπεται ότι:

  $$\det(A^T A) = \det(I_n) = 1 ,$$

  $$\det(A^T) \det(A) = 1 ,$$

  $$\det(A) = \pm 1$$

Γραμμική απεικόνιση

Η **γραμμική απεικόνιση** (ή γραμμικός μετασχηματισμός) είναι μια μαθηματική συνάρτηση που μετασχηματίζει διανύσματα από έναν διανυσματικό χώρο σε έναν άλλον, με τρόπο που διατηρεί δύο βασικές ιδιότητες:

1. **Γραμμικότητα ως προς την πρόσθεση**: Αν έχουμε δύο διανύσματα $\mathbf{v_1}$ και $\mathbf{v_2}$, τότε η γραμμική απεικόνιση **T** ικανοποιεί την εξίσωση: $\mathbf{T(v_1+v_2)=T(v_1)+T(v_2)}$
2. **Γραμμικότητα ως προς τον πολλαπλασιασμό με βαθμωτό αριθμό**: Αν έχουμε ένα διάνυσμα **v** και έναν αριθμό **c**, τότε ισχύει:$\mathbf{T(cv)=cT(v)}$

Αυτές οι δύο ιδιότητες καθορίζουν τη γραμμικότητα.

Οι πίνακες (μήτρες) συχνά χρησιμοποιούνται για να περιγράψουν τέτοιους γραμμικούς μετασχηματισμούς, και αυτό ακριβώς κάνει η εξίσωση **Ax=b**, όπου <u>η μήτρα **A** λειτουργεί ως η γραμμική απεικόνιση που μετασχηματίζει το διάνυσμα **x** στο αποτέλεσμα **b.**</u>

Σύστημα Γραμμικών Εξισώσεων **Ax=b**

$\mathbf{A} \in R^{m \times n}$, γνωστός πίνακας,

$\mathbf{b} \in R^{m}$, διάνυσμα,

$\mathbf{x} \in R^{n}$, διάνυσμα με τις άγνωστες μεταβλητές

Αναλύεται σε:

$$
\begin{aligned}
A_{1,:}x &= b_1 \\
A_{2,:}x &= b_2 \\
&\cdots\cdots\cdots\cdots \\
A_{m,:}x &= b_m
\end{aligned}
$$

ή

$$
\begin{aligned}
A_{1,1}x_1 + A_{1,2}x_2 + \cdots + A_{1,n}x_n &= b_1 \\
A_{2,1}x_1 + A_{2,2}x_2 + \cdots + A_{2,n}x_n &= b_2 \\
&\cdots\cdots\cdots\cdots \\
A_{m,1}x_1 + A_{m,2}x_2 + \cdots + A_{m,n}x_n &= b_m
\end{aligned}
$$

Μπορεί να έχει καμία λύση, πολλές λύσεις, ακριβώς μία λύση (πολ/σμό με ανάστροφο)

Πίνακες από διαφορετικές οπτικές γωνίες

Υπάρχουν τρεις συμπληρωματικές προοπτικές για την προβολή πινάκων:

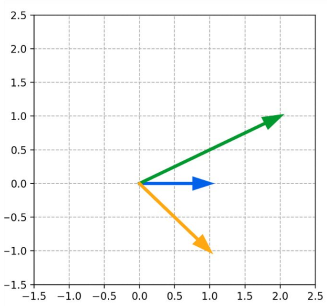
- **Προοπτική 1**: Ένας πίνακας ως πίνακας αριθμών
- **Προοπτική 2**: Ένας πίνακας ως λίστα διανυσμάτων (διανύσματα γραμμής και στήλης)
- **Προοπτική 3:** Ένας πίνακας ως συνάρτηση που αντιστοιχίζει διανύσματα από το ένα χώρο σε άλλο

Βλέποντας πίνακες μέσα από αυτές τις προοπτικές μπορούμε να αποκτήσουμε καλύτερη διαίσθηση για τους διανυσματικούς χώρους που προκαλούνται από τους πίνακες.

Διανύσματα στήλης

$$\begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & -1 \end{bmatrix}$$

Προοπτική 2

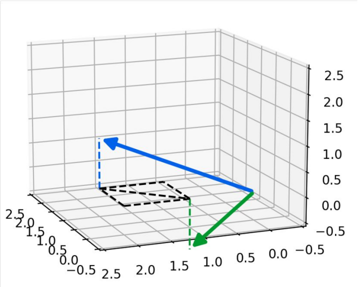


Προοπτική 3

Διανύσματα γραμμής

$$\begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & -1 \end{bmatrix}$$

Προοπτική 2



Προοπτική 3

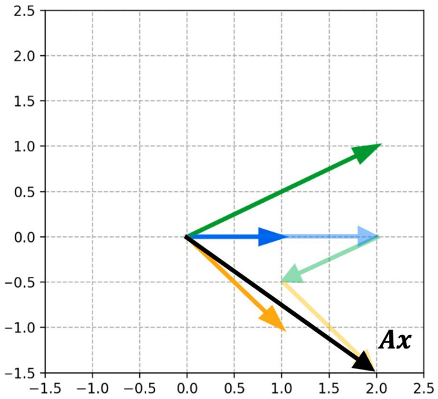
Προοπτική 3: Κατανόηση του column space

Για έναν δεδομένο πίνακα **A∈R$^{m\times n}$** , μπορούμε να δούμε αυτόν τον πίνακα ως συνάρτηση που αντιστοιχίζει διανύσματα από το **R^n** σε διανύσματα στο **R^m**

- Ένα διάνυσμα **x∈R^n** αντιστοιχίζεται στο διάνυσμα **b∈R^m** μέσω **Ax=b**
    - Ορίσουμε μια συνάρτηση **T: R^n→R^m ως: T(x):=Ax**
    - Αποδεικνύεται ότι το column space είναι απλώς το εύρος (range) αυτής της συνάρτησης T
        - Το εύρος αφορά τη γραμμική απεικόνιση που πραγματοποιεί ο πίνακας A, και ουσιαστικά είναι ο χώρος των εικόνων του A στο σύνολο των πιθανών εξόδων του

$$\underbrace{\begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & -1 \end{bmatrix}}_{\boldsymbol{A}} \underbrace{\begin{bmatrix} 2 \\ -0.5 \\ 1 \end{bmatrix}}_{\boldsymbol{x}} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} 2 + \begin{bmatrix} 2 \\ 1 \end{bmatrix} (-0.5) + \begin{bmatrix} 1 \\ -1 \end{bmatrix} 1 = \underbrace{\begin{bmatrix} 2 \\ -1.5 \end{bmatrix}}_{\boldsymbol{Ax}}$$

Πολλαπλασιασμός πίνακα (A) με διανύσμα (x): πράξη λήψης ενός γραμμικού συνδυασμού των στηλών του **A** χρησιμοποιώντας τους συντελεστές του **x** ως συντελεστές του γραμμικού συνδυασμού



Το column space είναι ολόκληρο το **R^2** αφού μπορούμε να σχηματίσουμε οποιοδήποτε δισδιάστατο διάνυσμα χρησιμοποιώντας έναν γραμμικό συνδυασμό αυτών των τριών διανυσμάτων

https://mbernste.github.io/posts/matrixspaces/

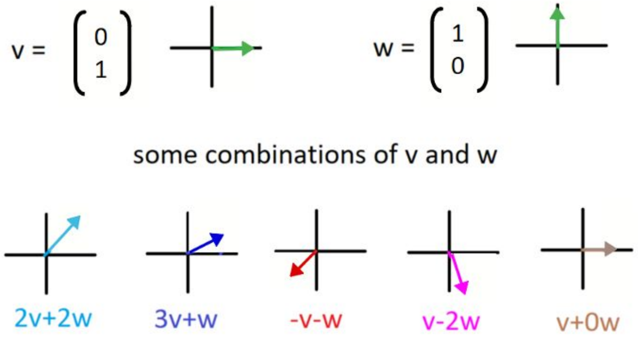
Σύστημα Γραμμικών Εξισώσεων **Ax=b**

Σκέψη :

Οι στήλες του A ορίζουν διαφορετικές κατευθύνσεις που μπορούμε να “κινηθούμε” στο χώρο από το αρχικό σημείο 0 για να προσεγγίσουμε το b.

$$Ax = \sum_i x_i A_{:,i} \rightarrow \text{Γραμμικός συνδυασμός}$$

Το **εύρος (span)** ενός είναι **το σύνολο όλων των** ιπορούν να ληφθούν **με γραμμι** **ων αρχικών διανυσμάτων.**



Σύστημα Γραμμικών Εξισώσεων **Ax=b**

Έχει λύση η **Ax = b**;

Θα ελέγχουμε αν το **b** είναι στο **εύρος των στηλών** του **A**, δηλαδή αν το **b** εκφραστεί ως γραμμικός συνδυασμός των στηλών του **A**.

Τρόποι Ελέγχου

1. Μέθοδος ελέγχου μέσω κατάταξης (rank)
2. Λύση μέσω απαλοιφής Gauss
3. Λύση μέσω Ορθογώνιας Προβολής (π.χ. Least Squares)
4. Επίλυση μέσω Singular Value Decomposition (SVD)

Για να μπορεί το σύστημα **Ax = b** να έχει λύση για όλες τις τιμές του $\mathbf{b} \in \Re_m$, απαιτούμε ο χώρος των στηλών του **A** να είναι ο $\Re_m$.

Πρέπει ο **A** να έχει ακριβώς **m** γραμμικές ανεξάρτητες στήλες, όχι τουλάχιστον **m**.

Σύστημα Γραμμικών Εξισώσεων

$$\mathbf{Ax = b}$$

Προκειμένου ο πίνακας **A** να έχει αντίστροφο, πρέπει επιπλέον να διασφαλίσουμε ότι η εξίσωση έχει το πολύ μία λύση για κάθε τιμή του **b**. (Διαφορετικά, υπάρχουν περισσότεροι από έναν τρόποι παραμετροποίησης κάθε λύσης)

$$Ax = b$$
$$A^{-1}Ax = A^{-1}b$$
$$I_n x = A^{-1}b$$
$$x = A^{-1}b$$

Για να έχει αντίστροφο ο πίνακας A θα πρέπει:

- **να είναι τετράγωνος, δηλαδή, απαιτούμε ότι m = n και**
- **όλες οι στήλες πρέπει να είναι γραμμικώς ανεξάρτητες**

Τριγωνική παραγοντοποίηση A=LU

Αν δεν απαιτούνται εναλλαγές γραμμών, ο αρχικός πίνακας μπορεί να γραφεί ως γινόμενο A=LU

L : κάτω τριγωνικός

U : άνω τριγωνικός

$$L = \begin{bmatrix} 1 & 0 & .. & 0 \\ \lambda_{2,1} & 1 & .. & 0 \\ .. & .. & .. & .. \\ \lambda_{n,1} & \lambda_{n,2} & .. & 1 \end{bmatrix} \qquad U = \begin{bmatrix} u_{1,1} & u_{1,2} & \dots & u_{1,n} \\ 0 & u_{2,2} & \dots & u_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & u_{m,n} \end{bmatrix}$$

- Εμφανίζεται **μετά τη διαδοχική απαλοιφή** και **πριν την ανάδρομη αντικατάσταση**.
- Τα **διαγώνια στοιχεία του είναι οι οδηγοί**

π.χ.

$$A = \begin{bmatrix} 3 & 1 \\ 4 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \frac{4}{3} & 1 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 0 & \frac{2}{3} \end{bmatrix} = LU$$

$$A = \begin{bmatrix} -5 & 1 & -3 & 4 \\ 8 & -7 & 3 & 2 \\ -3 & -6 & -1 & -1 \\ 0 & 0 & 3 & 9 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -\frac{8}{5} & 1 & 0 & 0 \\ \frac{3}{5} & \frac{33}{27} & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} -5 & 1 & -3 & 4 \\ 0 & -\frac{27}{5} & -\frac{9}{5} & \frac{42}{5} \\ 0 & 0 & 3 & -\frac{41}{3} \\ 0 & 0 & 0 & \frac{68}{3} \end{bmatrix} = LU$$

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 7 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & 0 & 0 \end{bmatrix} = LU$$

Εύρεση A^{-1} : Μέθοδος των Gauss-Jordan

Έστω η εξίσωση : $AA^{-1}=I$.

Εάν θεωρηθεί στήλη προς στήλη αυτή η εξίσωση προσδιορίζει τις στήλες του A^{-1}

$$Ax_1 = e_1, Ax_2 = e_2, Ax_3 = e_3$$

όπου

$$\begin{bmatrix} e_1 & e_2 & e_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} A & e_1 & e_2 & e_3 \end{bmatrix} = \begin{bmatrix} U & L \end{bmatrix}$$

$$\begin{bmatrix} U & L \end{bmatrix} = \begin{bmatrix} I & A^{-1} \end{bmatrix}$$

Εύρεση A^{-1} : Μέθοδος των Gauss-Jordan

$$A=\begin{pmatrix}9&3&4\\4&3&4\\1&1&1\end{pmatrix}$$

$$(A \mid I_3)=\left(\begin{array}{ccc|ccc}9&3&4&1&0&0\\4&3&4&0&1&0\\1&1&1&0&0&1\end{array}\right)=\left(\begin{array}{ccc|ccc}9&3&4&1&0&0\\0&5/3&20/9&-4/9&1&0\\0&2/3&5/9&-1/9&0&1\end{array}\right);$$

$$[\,A \quad e_1 \quad e_2 \quad e_3\,]=[\,U \quad L\,]$$
$$[\,U \quad L\,]=[\,I \quad A^{-1}\,]$$

$$=\left(\begin{array}{ccc|ccc}9&0&0&9/5&-9/5&0\\0&5/3&20/9&-4/9&1&0\\0&0&-1/3&1/15&-2/5&1\end{array}\right)$$

$$=\left(\begin{array}{ccc|ccc}9&0&0&9/5&-9/5&0\\0&5/3&0&0&-5/3&20/3\\0&0&-1/3&1/15&-2/5&1\end{array}\right)$$

Διαιρούμε όλα τα στοιχεία της i-γραμμής του πίνακα εκ δεξιών της διακεκκομένης γραμμής με το μη μηδενικό στοιχείο της i- γραμμής του διαγώνιου πίνακα και παίρνουμε:

$$A^{-1}=\begin{pmatrix}1/5&-1/5&0\\0&-1&4\\-1/5&6/5&-3\end{pmatrix}.$$

Παραγοντοποίηση A=LDU

- Ο **πίνακας L παραμένει ίδιος** όπως στην A=LU
- Ο **πίνακας D** είναι διαγώνιος και περιέχει **την διαγώνιο του πίνακα U** της παραγοντοποίησης LU (στοιχεία οδηγών):
- Ο **νέος άνω τριγωνικός πίνακας U** προκύπτει από τον πίνακα U της παραγοντοποίησης LU **διαιρώντας κάθε στοιχείο του με το στοιχείο της διαγωνίου** (δηλ. τον οδηγό) της ίδιας γραμμής

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 7 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1/2 \end{bmatrix} = LDU$$

$$A= \begin{pmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 7 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 0 & 3 & 6 \\ 0 & 0 & 4 \end{pmatrix} =LU$$

Νόρμες (Norms)

Η **νόρμα** είναι μια συνάρτηση που αναθέτει έναν θετικό αριθμό (ή μηδέν) σε έναν **διανυσματικό χώρο** ή έναν **πίνακα**, με στόχο να μετρήσει το "μέγεθος" ή το "μήκος" ενός διανύσματος ή πίνακα.

Ουσιαστικά, η νόρμα είναι ένας τρόπος μέτρησης της απόστασης από το μηδέν στο διάστημα που εξετάζουμε.

$$L^p = \|x\|_p = \left(\sum |x_i|^p\right)^{1/p} \qquad p \subset \Re, p \geq 1$$

Η νόρμα είναι μία συνάρτηση f που ικανοποιεί τις ακόλουθες ιδιότητες:

1. **Μη αρνητικότητα**: η νόρμα είναι πάντα θετική εκτός από την περίπτωση του μηδενικού διανύσματος, που έχει νόρμα μηδέν.
2. **Ανισότητα τριγώνου**: η νόρμα του αθροίσματος δύο διανυσμάτων είναι μικρότερη ή ίση από το άθροισμα των νόρμων τους (παρόμοια με την ανισότητα τριγώνου στη γεωμετρία).
3. **Ομογένεια (homogeneity)**: η νόρμα πολλαπλασιάζεται απόλυτα με το μέτρο του αριθμού α.

$$f(\boldsymbol{x}) = 0 \Rightarrow \boldsymbol{x} = \boldsymbol{0}$$

$$f(\boldsymbol{x}+\boldsymbol{y}) \leq f(\boldsymbol{x}) + f(\boldsymbol{y}) \text{ (the \textbf{triangle inequality})}$$

$$\forall \alpha \in \mathbb{R}, f(\alpha \boldsymbol{x}) = |\alpha| f(\boldsymbol{x})$$

L^2 νόρμα (Ευκλείδια νόρμα)

- Για p=2 → η ευκλείδεια απόσταση από την αρχή μέχρι το σημείο που προσδιορίζεται από το x.
- Δηλώνεται συχνά ως || x ||, με το δείκτη 2 να παραλείπεται.
- Συνήθως υπολογίζουμε το τετράγωνο της L^2 νόρμας, απλά ως $x^{\mathrm{T}}x$
- Υπολογιστικά, συχνά, το τετράγωνο της L^2 νόρμα μπορεί να αυξάνεται πολύ αργά κοντά το σημείο που προσδιορίζεται από το x
  → απαγορευτικό υπολογιστικά

L^1 νόρμα

- Συχνά είναι σημαντικό να γίνεται διάκριση μεταξύ στοιχείων που είναι ακριβώς μηδενικά και αυτών που είναι μικρά αλλά μη μηδενικά.

$$\|x\|_1 = \sum_i |x_i|$$

- Κάθε φορά που ένα στοιχείο του **x** μετακινείται κατά **e** μακριά από το 0, η $\mathbf{L^1}$ αυξάνεται κατά **e**

- Υποκαθιστά το πλήθος των μη μηδενικών τιμών.

Μερικές φορές μετράμε το μέγεθος του διανύσματος υπολογίζοντας τον αριθμό των μη-μηδενικών στοιχείων (Λανθασμένη ορολογία το νόρμα L^0).

L^{∞} νόρμα (Max νόρμα)

Υπολογίζει την απόλυτη τιμή του μεγαλύτερου στοιχείου του διανύσματος x

$$\|x\|_{\infty} = \max_{i} |x_i|$$

Υπολογισμός μεγέθους πίνακα

Νόρμα Frobenius

$$\| A \|_F = \left(\sum_{i,j} | a_{ij} |^2\right)^{1/2}$$

Είναι το ανάλογο της L^2 ενός διανύσματος.

Εσωτερικό γινόμενο με χρήση νόρμας

$$x^T y = \|x\|_2 \|y\|_2 cos\theta$$, όπου θ η γωνία μεταξύ x και y

Διαγώνιοι πίνακες

Ένας πίνακας D είναι διαγώνιος εάν και μόνο εάν $D_{i,j} = 0$ για όλα τα $i \neq j$

$diag(v)$: τετραγωνικός διαγώνιος πίνακας,
το διάνυσμα **u** περιέχει τις τιμές της διαγωνίου

Ο πολ/σμός του με άλλο πίνακα είναι υπολογιστικά αποδοτικός.

$$diag(v)x = v \odot x$$

Αντίστροφος : $\exists$ αν τα στοιχεία της διαγωνίου είναι μη μηδενικά

$$diag(v)^{-1} = diag([1/v_1, \ldots, 1/v_n])$$

- Σε πολλές περιπτώσεις, αλγόριθμοι μηχανικής μάθησης περιορίζουν πίνακες να είναι διαγώνιοι, κερδίζοντας υπολογιστικό κόστος και χρόνο.
- Είναι δυνατή η κατασκευή διαγώνιου πίνακα

Ιδιοτιμές και ιδιοδιανύσματα

Επίλυση γραμμικών συστημάτων

Επινοήθηκαν για επίλυση διαφορικών εξισώσεων :

Η λύση θα είναι της μορφής :

$$\left.\begin{array}{l} \dfrac{dy}{dt} = Ay \\ y(t) = e^{\lambda t} x \end{array}\right] \Rightarrow \lambda e^{\lambda t} x = \mathrm{A} e^{\lambda t} x \Rightarrow \mathrm{A} x = \underset{\text{Ιδιοτιμή}}{\lambda} \underset{\text{Ιδιοδιάνυσμα}}{x}$$

- Ο αριθμός λ είναι ιδιοτιμή του A όταν και μόνον όταν ισχύει $det(A - \lambda I) = 0$
- Αυτή είναι η χαρακτηριστική εξίσωση και σε κάθε λύση της λ αντιστοιχεί ένα ιδιοδιάνυσμα x:

$$(A - \lambda I)x = 0 \quad \text{ή} \quad Ax = \lambda x$$

Βήματα επίλυσης του προβλήματος ιδιοτιμών

1. Υπολογισμός της ορίζουσας του
2. Εύρεση των ριζών αυτού του πολυωνύμου $\quad A - \lambda\mathrm{I} \quad \rightarrow$ ιδιοτιμές
3. Για κάθε ιδιοτιμή, επίλυση του συστήματος $\quad (A - \lambda\mathrm{I})x = 0 \rightarrow$ ιδιοδιανύσματα

π.χ. Διαφορική εξίσωση

$$\frac{dy_1}{dt} = 4y_1 - 5y_2$$
$$\frac{dy_2}{dt} = 2y_1 - 3y_2$$
$$t = 0 : y_1 = 8, y_2 = 5$$

$$A = \begin{bmatrix} 4 & -5 \\ 2 & -3 \end{bmatrix}$$

1., 2.

$$|A - \lambda I| = \begin{bmatrix} 4 & -5 \\ 2 & -3 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 4-\lambda & -5-0 \\ 2-0 & -3-\lambda \end{bmatrix} = (4-\lambda)(-3-\lambda) + 10 = -(4-\lambda)(3+\lambda) + 10 = (\lambda-4)(\lambda+3) - 10 = \lambda^2 - 4\lambda + 3\lambda - 12 + 10 = \lambda^2 - \lambda - 2 \rightarrow \lambda_1 = -1, \lambda_2 = 2$$

3.

$$(A - \lambda_1 I)x = \begin{bmatrix} 5 & -5 \\ 2 & -2 \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \rightarrow x_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$(A - \lambda_2 I)x = \begin{bmatrix} 2 & -5 \\ 2 & -5 \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \rightarrow x_2 = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$$

Ειδικές λύσεις :

$$u = e^{\lambda_1 t} x_1 = e^{-t} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$u = e^{\lambda_2 t} x_2 = e^{2t} \begin{bmatrix} 5 \\ 2 \end{bmatrix}$$

Γενική λύση: $u = c_1 e^{\lambda_1 t} x_1 + c_2 e^{\lambda_2 t} x_2$

Χρήση αρχικών συνθηκών $\rightarrow$ $8 = c_1 + 5c_2$
$5 = c_1 + 2c_2$

$$u = 3e^{-t} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + e^{2t} \begin{bmatrix} 5 \\ 2 \end{bmatrix}$$

Ιδιοδιανύσματα (eigenvectors)

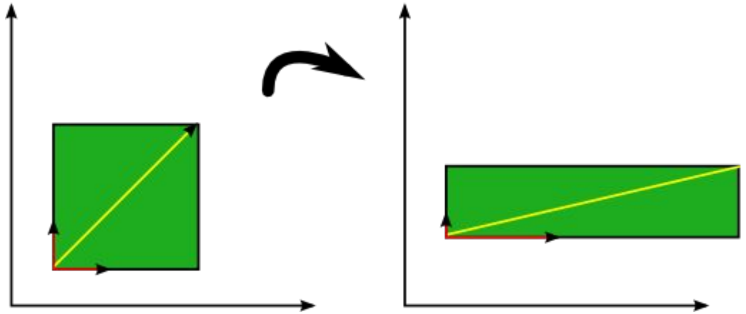
Ένα **ιδιοδιάνυσμα** ενός γραμμικού μετασχηματισμού είναι ένα μη μηδενικό διάνυσμα που, όταν εφαρμοστεί πάνω του ο γραμμικός μετασχηματισμός (π.χ., μέσω ενός πίνακα A), δεν αλλάζει κατεύθυνση.

→ δηλαδή, η δράση του μετασχηματισμού στο ιδιοδιάνυσμα οδηγεί απλά σε πολλαπλασιασμό του με έναν σταθερό αριθμό.

Για έναν πίνακα A και ένα ιδιοδιάνυσμα v, ισχύει: **Av=λv**

όπου:

- A είναι ο πίνακας που εκφράζει τον γραμμικό μετασχηματισμό.
- v είναι το ιδιοδιάνυσμα.
- λ είναι μια σταθερά, η **ιδιοτιμή** (eigenvalue), που δείχνει πόσο "κλιμακώνεται" το ιδιοδιάνυσμα.



Τα ιδιοδιανύσματα (κόκκινο) δεν αλλάζουν κατεύθυνση όταν εφαρμόζεται σε αυτά ένας γραμμικός μετασχηματισμός (π.χ. κλιμάκωση).

- Μπορεί να αλλάξουν μέγεθος, αλλά η κατεύθυνσή τους παραμένει ίδια.

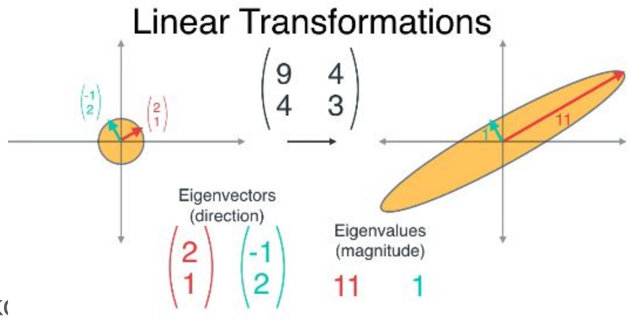
Άλλα διανύσματα (κίτρινο), που δεν είναι ιδιοδιανύσματα αλλάζουν.

Ιδιοτιμές (Eigenvalues)

Οι **ιδιοτιμές** είναι οι αντίστοιχοι κλιμακωτοί παράγοντες που συνδέονται με κάθε ιδιοδιάνυσμα.

Δηλαδή, όταν εφαρμόζεται ο πίνακας **A** σε ένα ιδιοδιάνυσμα **v**, το αποτέλεσμα είναι να "τεντώνεται" ή να "συρρικνώνεται" το **v** κατά έναν παράγοντα **λ**, αλλά η κατεύθυνσή του παραμένει η ίδια.

- Η ιδιοτιμή **λ** μετράει την **ποσότητα** κατά την οποία το αντίστοιχο ιδιοδιάνυσμα **v** τεντώνεται ή συρρικνώνεται από τον γραμμικό μετασχηματισμό.
    - αν λ>1, το διάνυσμα μεγαλώνει,
    - αν 0<λ<1, συρρικνώνεται,
    - αν λ<0, το διάνυσμα αντιστρέφει κατεύθυνση.



Εφαρμογές

- **Μείωση διαστάσεων (PCA)**
Οι κύριες συνιστώσες είναι τα ιδιοδιανύσματα ενός πίνακα συνδιακύμανσης, κα
κύριας συνιστώσας.
- **Δυναμικά Συστήματα**:
Τα ιδιοδιανύσματα και οι ιδιοτιμές βοηθούν στην ανάλυση της συμπεριφοράς ενός συστήματος, π.χ., αν ένα σύστημα θα συγκλίνει σε μια σταθερή κατάσταση.
- **Γραμμικοί Μετασχηματισμοί**
Οι ιδιοτιμές χρησιμοποιούνται για την εύρεση λύσεων και την κατανόηση της δομής του πίνακα.

Ιδιοτιμές και ιδιοδιανύσματα

Εξίσωση-κλειδί: $Ax = \lambda x$

- Οι **ιδιοτιμές** παριστάνουν το **βηματισμό στο χρόνο** (magnitude)

- Τα **ιδιοδιανύσματα** παριστάνουν τις “**κανονικές καταστάσεις” του συστήματος** και επιδρούν ανεξάρτητα το ένα από το άλλο (direction)

- Μπορούμε να παρακολουθήσουμε τη <u>συμπεριφορά κάθε ιδιοδιανύσματος</u> και να συνδυάσουμε αυτές τις κανονικές καταστάσεις για να βρούμε <u>μια λύση</u>

  → Διαγωνοποίηση

Διαγωνοποίηση

$$A = V\Lambda V^{-1}$$

$$\left.\begin{aligned} \frac{dy}{dt} &= Ay \\ y(t) &= e^{\lambda t}x \end{aligned}\right\} \Rightarrow \lambda e^{\lambda t}x = \mathrm{A}e^{\lambda t}x \Rightarrow \mathrm{A}x = \lambda x$$

$$\Rightarrow A\begin{bmatrix} x_1 & x_2 \end{bmatrix} = \begin{bmatrix} \lambda_1 x_1 & \lambda_2 x_2 \end{bmatrix}$$

Eigenvector Matrix

Eigenvalue Matrix

$$\Rightarrow A\begin{bmatrix} x_1 & x_2 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

$$\Rightarrow AV = V\Lambda$$

$$\Rightarrow AVV^{-1} = V\Lambda V^{-1}$$

$$\Rightarrow AI = V\Lambda V^{-1}$$

$$\Rightarrow \boxed{A = V\Lambda V^{-1}}$$

A^2 = ;

Ο πίνακας ιδιοτιμών υψωμένος στο τετράγωνο

$$A^2 = V\Lambda V^{-1}V\Lambda V^{-1} = V\Lambda I\Lambda V^{-1} = V\Lambda^2 V^{-1}$$

A^n = ;

Ίδιος πίνακας ιδιοδιανυσμάτων όπως ο Α

$$\boxed{A^n = V\Lambda^n V^{-1}}$$

Ο πίνακας ιδιοτιμών υψωμένος στην n

Πίνακες, ιδιοτιμές, ιδιοδιανύσματα

Συμμετρικός πίνακας $S = S^T$ $\rightarrow$ πραγματικές ιδιοτιμές

$\rightarrow$ ορθογώνια ιδιοδιανύσματα

Ανάστροφος πίνακας $\mathrm{A}^{\mathrm{T}} = -\mathrm{A}$ $\rightarrow$ φανταστικές ιδιοτιμές

$\rightarrow$ ορθογώνια μιγαδικά ιδιοδιανύσματα

Ορθογώνιος πίνακας $Q^{\mathrm{T}}Q = I$ $\rightarrow$ για όλες τις ιδιοτιμές ισχύει: $|\lambda| = 1$

$\rightarrow$ ορθογώνια μιγαδικά ιδιοδιανύσματα

Ένας πίνακας ονομάζεται **ιδιάζων (singular)** ανν οποιαδήποτε από τις ιδιοτιμές του είναι μηδέν.

Πίνακες, ιδιοτιμές, ιδιοδιανύσματα

Ορισμένος πίνακας

- Ένας **συμμετρικός n×n** πίνακας ονομάζεται **θετικά** (αρνητικά) ορισμένος, εάν για όλα τα μη μηδενικά διανύσματα $x \in R^n$ το $Q(x) = x^T A x$ παίρνει μόνο θετικές τιμές (αρνητικές τιμές).

- Εάν ένας **συμμετρικός nxn** πίνακας παίρνει μόνο θετικές (αρνητικές) ονομάζεται **γνησίως-θετικός** (αρνητικός)

$$x^T A x = 0 \Rightarrow x = 0$$

- Εάν ένας **συμμετρικός nxn** πίνακας παίρνει θετικές (αρνητικές) ή μηδενικές ονομάζεται **ημι-θετικός** (αρνητικός)

$$\forall x, x = x^T A x \geq 0$$

- Εάν ένας **συμμετρικός nxn** δεν είναι ούτε θετικός ούτε αρνητικός τότε ο πίνακας είναι **αόριστος**.

Παραγοντοποίηση

$A=LU$ (Απαλοιφή)

$A=QR$ (Gram-Schmidt)

$S=Q\Lambda Q^T$ (Symmetric: $\begin{bmatrix}\sigma_1 & \cdots & \sigma_n\end{bmatrix}\begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{bmatrix}\begin{bmatrix}\sigma_1^T \\ \vdots \\ \sigma_n^T\end{bmatrix}$)

$A=X\Lambda X^{-1}$

$A=U\Sigma V^T$ (Ορθογώνιος x Διαγώνιος x Ορθογώνιος)

Trace Operator

Δίνει το άθροισμα όλων των διαγωνίων τιμών ενός πίνακα $Tr(\mathbf{A}) = \sum_i A_{i,i}$

Frobenius νόρμα του A: $\|\mathbf{A}\|_F = \sqrt{Tr(AA^T)}$

Παράδειγμα

$$Tr\begin{bmatrix} a_{00} & a_{01} & a_{02} \\ a_{10} & a_{11} & a_{12} \\ a_{20} & a_{21} & a_{22} \end{bmatrix} = a_{00} + a_{11} + a_{22}$$

<u>ΙΔΙΟΤΗΤΕΣ</u>

$$Tr(A) = Tr(A^T)$$

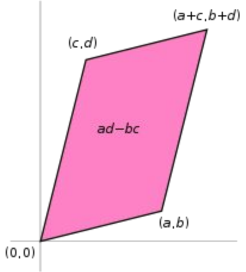
$Tr(ABC) = Tr(CAB) = Tr(BCA)$, αν το επιτρέπουν οι διαστάσεις των πινάκων ή

$$Tr(\prod_{i=1}^{n} F^{(i)}) = Tr(F^{(n)} \prod_{i=1}^{n} F^{(i)})$$

Ένα βαθμωτό είναι το δικό του ίχνος: $a = Tr(a)$

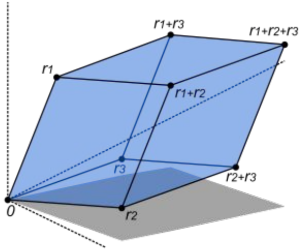
Ορίζουσα (Determinant)

det(A) : αντιστοιχεί πίνακα σε βαθμωτό



- Ισούται με το **γινόμενο όλων των ιδιοτιμών του πίνακα**
- Γεωμετρική ερμηνεία
  - Η απόλυτη τιμή της ορίζουσας δίνει την κλίμακα με την οποία το εμβαδόν ή ο όγκος (ή μιας μεγαλύτερης διάστασης αναλογία) πολλαπλασιάζεται με τον σχετικό γραμμικό μετασχηματισμό,
  - το πρόσημό της δείχνει αν ο μετασχηματισμός διατηρεί τον προσανατολισμό.

Συνοπτικά, η ορίζουσα παρέχει γεωμετρική πληροφορία σχετικά με τη διάταξη και την κατεύθυνση των διανυσμάτων στο επίπεδο.



Ορίζουσα Πίνακα: Γεωμετρική ερμηνεία

Έστω ότι έχουμε 2 διανύσματα στο επίπεδο **v**=[a c] και **w**=[b d].

Ο πίνακας που σχηματίζεται από αυτά τα δύο διανύσματα είναι: $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$

Η ορίζουσα det(A)=ad-cb, είναι επίσης η περιοχή του παραλληλογράμμου που σχηματίζεται από τα δύο διανύσματα.

- Αν η ορίζουσα είναι θετική, τότε τα δύο διανύσματα δείχνουν προς ίδια κατεύθυνση και η περιοχή του παραλληλογράμμου είναι θετική.
- Αν είναι αρνητική, τα δύο διανύσματα δείχνουν προς αντίθετες κατευθύνσεις και η περιοχή είναι αρνητική.
- Αν η ορίζουσα είναι μηδέν, τότε τα δύο διανύσματα είναι γραμμικά εξαρτημένα, και η περιοχή του παραλληλογράμμου είναι μηδενική.

π.χ. $\mathbf{A}_{2\times 2}$, det(A)=−2 : όταν εφαρμόζεται στην περιοχή ενός επιπέδου με πεπερασμένο εμβαδόν, θα μετασχηματιστεί σε μια περιοχή με το διπλάσιο εμβαδόν, ενώ αντιστρέφει τον προσανατολισμό της.

det(A)=0 : ο χώρος συστέλλεται

det(A)=1 : ο μετασχηματισμός διατηρεί τον όγκο

Πιθανότητες

Βασικές Έννοιες Πιθανοτήτων

- Η πιθανότητα εκφράζει την **αβεβαιότητα ενός γεγονότος**.
- Χρησιμοποιείται για την **εκτίμηση** του πόσο πιθανό είναι να συμβεί ένα γεγονός.

Βασικές έννοιες

- Τυχαίες μεταβλητές
- Κατανομές πιθανότητας
- Συναρτήσεις μάζας και πυκνότητας πιθανότητας

Τυχαίες Μεταβλητές

Μια τυχαία μεταβλητή είναι μια μεταβλητή που μπορεί **να πάρει διαφορετικές τιμές** *σύμφωνα με κάποιες πιθανότητες*.

Χρησιμοποιούνται για **την περιγραφή τυχαίων φαινομένων**.

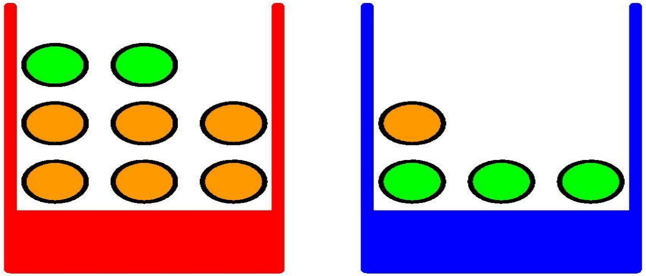
Οι τυχαίες μεταβλητές χωρίζονται σε δύο κατηγορίες:

- **Διακριτές Τυχαίες Μεταβλητές**: συγκεκριμένες και μετρήσιμες τιμές
  - ο αριθμός των παιδιών σε μια οικογένεια (0, 1, 2, ...),
  - οι ρίψεις ενός ζαριού (1, 2, 3, 4, 5, 6).

- **Συνεχείς Τυχαίες Μεταβλητές:** οποιαδήποτε τιμή σε ένα συνεχές διάστημα.
  - το ύψος ενός ατόμου (μπορεί να είναι 1.75m, 1.76m, 1.761m κ.ο.κ.),
  - η θερμοκρασία της ατμόσφαιρας.

Probability Theory - προσέγγιση

Apples and Oranges



Πείραμα

1. Τυχαία επιλογή δοχείου
2. Τυχαία επιλογή φρούτου
3. Επανατοποθέτηση φρούτου

Βασική έννοια στην αναγνώριση προτύπων:

Αβεβαιότητα

→ θόρυβος στις μετρήσεις

→ πεπερασμένο σύνολο δεδομένων

Θεωρία πιθανοτήτων

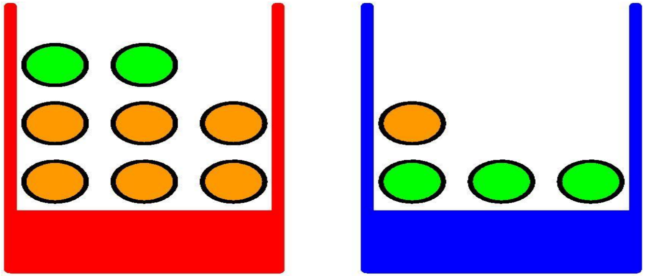
→ ποσοτικοποίηση της αβεβαιότητας

→ χειρισμός της αβεβαιότητας

Θεωρία πιθανοτήτων + θεωρία αποφάσεων → Βέλτιστες Προβλέψεις

Probability Theory - προσέγγιση

Apples and Oranges



→ Επανάληψη του πειράματος πολλές φορές

→ Έστω 40% επιλέχθηκε το red Box και 60% το blue Box

> Τυχαία μεταβλητή: B = { r , b}

> Τυχαία μεταβλητή: F={ a , o}

ρ(B = r) = 4/10

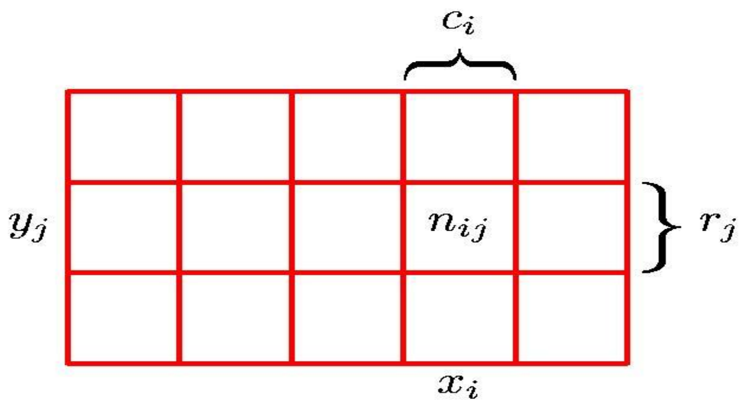
p(B = β)=6/10

Πιθανά ερωτήματα:

- Ποια είναι η συνολική πιθανότητα να επιλέξουμε μήλο;
- Αν έχουμε επιλέξει ένα πορτοκάλι, ποια είναι η πιθανότητα να είναι από το μπλε δοχείο;

Κανόνες αθροίσματος και γινομένου πιθανοτήτων

Probability Theory



X = x_i, i={1..M} (**Boxes**)

Y = $y_{j,}$ j={1..L} (**Fruits**)

N: συνολικές δοκιμές

n_{ij}= πλήθος δοκιμών όπου $X = x_i$ *και* $Y = y_j$

c_i= πλήθος δοκιμών όπου $X = x_i$ *ανεξάρτητα του Y*

r_j= πλήθος δοκιμών όπου $Y = Y_j$ *ανεξάρτητα του X*

Marginal Probability
Οριακή Πιθανότητα

$$p(X = x_i) = \frac{c_i}{N}.$$

Joint Probability
Από κοινού πιθανότητα

$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$$

Conditional Probability
Υπό Συνθήκη Πιθανότητα

$$p(Y = y_j | X = x_i) = \frac{n_{ij}}{c_i}$$

Probability Theory

X = x_i, i={1..M} (**Boxes**)

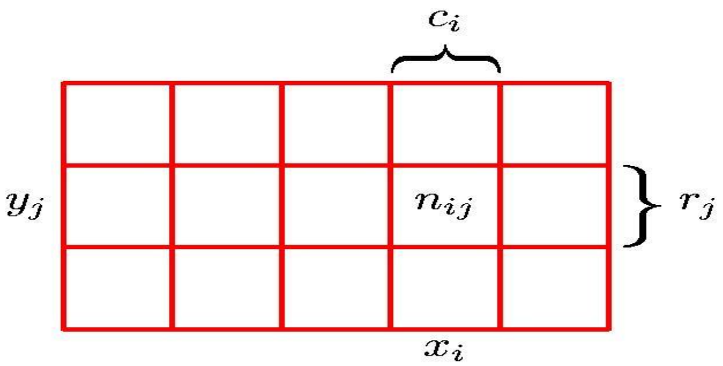
Y = y_j, j={1..L} (**Fruits**)

N: συνολικές δοκιμές

n_{ij}= πλήθος δοκιμών όπου $X = x_i$ *και* $Y = y_j$

c_i= πλήθος δοκιμών όπου $X = x_i$ *ανεξάρτητα του Y*

r_j= πλήθος δοκιμών όπου $Y = Y_j$ *ανεξάρτητα του X*



Sum Rule

$$p(X = x_i) = \frac{c_i}{N} = \frac{1}{N}\sum_{j=1}^{L} n_{ij}$$

$$= \sum_{j=1}^{L} p(X = x_i, Y = y_j)$$

Product Rule

$$\begin{aligned} p(X = x_i, Y = y_j) &= \frac{n_{ij}}{N} = \frac{n_{ij}}{c_i} \cdot \frac{c_i}{N} \\ &= p(Y = y_j | X = x_i) p(X = x_i) \end{aligned}$$

The Rules of Probability

Sum Rule $$p(X) = \sum_Y p(X, Y)$$

Product Rule $$p(X, Y) = p(Y|X)p(X)$$

Bayes' Theorem

$$\boxed{p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)}}$$

όπου:

$$p(X) = \sum_Y p(X|Y)p(Y)$$

posterior ∝ likelihood × prior

$$p(X,Y) = p(Y|X)p(X)$$

όπου: $$p(X) = \sum_Y p(X,Y)$$

X = x_i, i={1..M} (**Boxes**)

Y = y_j, j={1..L} (**Fruits**)

Ποιο **κουτί** επιλέχθηκε **πριν** μάθουμε ποιο **φρούτο** επιλέχθηκε;

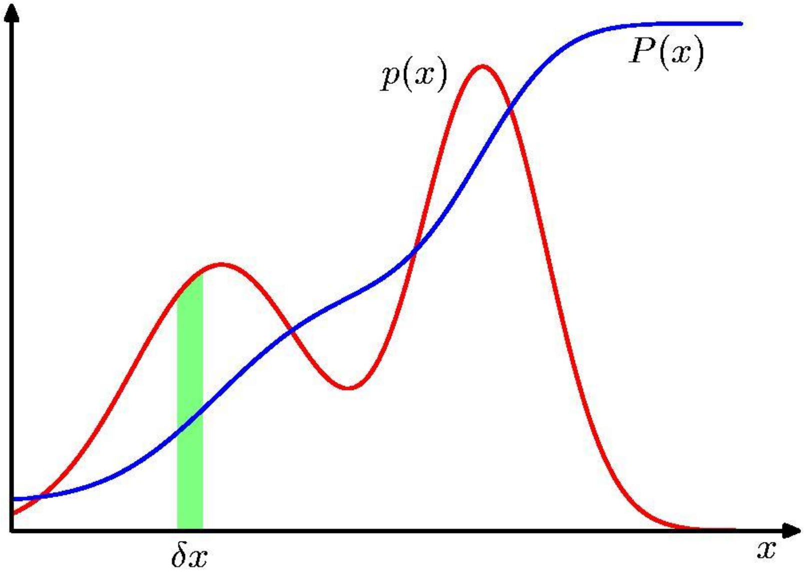
→ **prior:** η πιθανότητα είναι γνωστή πριν διαλέξουμε το φρούτο.

Μόλις διαλέξουμε φρούτο, έστω το πορτοκάλι, με το <u>θεώρημα του Bayes</u>, μπορούμε να υπολογίζουμε την **p(B,F)**

→ **posterior :** η πιθανότητα p(B,F) που λαμβάνουμε αφού έχουμε παρατηρήσει το F.

Probability Densities

Για διακριτές μεταβλητές → **Πυκνότητα Πιθανότητας**



$$p(x \in (a,b)) = \int_a^b p(x)\,\mathrm{d}x$$

$$P(z) = \int_{-\infty}^{z} p(x)\,\mathrm{d}x$$

$$p(x) \geqslant 0$$

$$\int_{-\infty}^{\infty} p(x)\,\mathrm{d}x = 1$$

Expectations

Προσδοκία: η μέση τιμή μιας συνάρτησης *f(x)* κάτω από κατανομή πιθανότητας *p(x)*

Διακριτή Κατανομή

$$\mathbb{E}[f] = \sum_x p(x) f(x)$$

Συνεχείς Μεταβλητές

$$\mathbb{E}[f] = \int p(x) f(x) \, \mathrm{d}x$$

Δεσμευμένη Προσδοκία

$$\mathbb{E}_x[f|y] = \sum_x p(x|y) f(x)$$

Conditional Expectation (discrete)

$$\mathbb{E}[f] \simeq \frac{1}{N} \sum_{n=1}^{N} f(x_n)$$

Approximate Expectation (discrete and continuous)

Variances and Covariances

$$\mathbb{E}[f] = \sum_x p(x) f(x)$$

Διακύμανση της *f(x)* : μέτρο μεταβλητότητας που υπάρχει στην *f(x) γύρω από τη μέση τιμή της E[f(x)}*

$$\mathrm{var}[f] = \mathbb{E}\left[\left(f(x) - \mathbb{E}[f(x)]\right)^2\right] = \mathbb{E}[f(x)^2] - \mathbb{E}[f(x)]^2$$

Συνδιακύμανση 2 τυχαίων μεταβλητών x,y: εκφράζει την έκταση που τα x και y μεταβάλλονται από κοινού

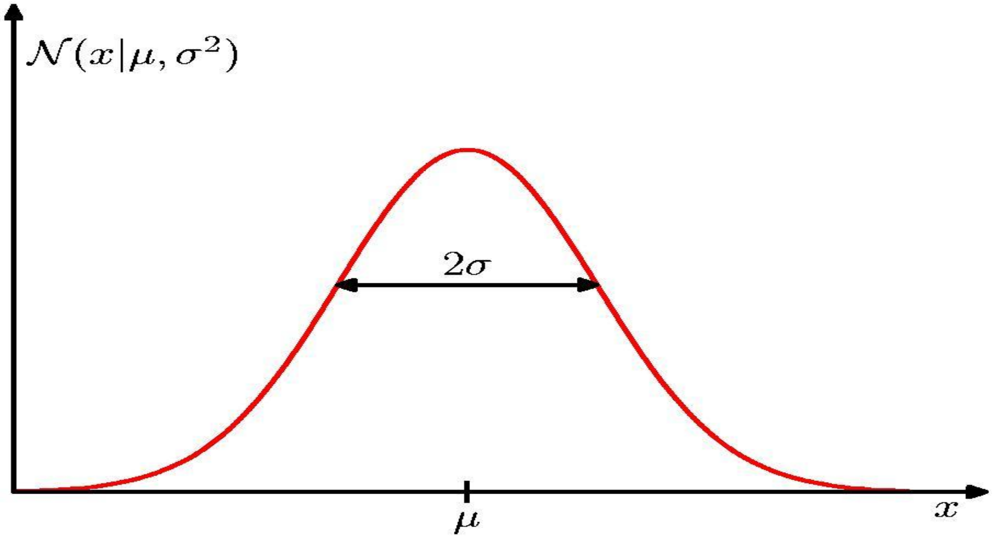
$$\begin{aligned} \mathrm{cov}[x, y] &= \mathbb{E}_{x,y}\left[\{x - \mathbb{E}[x]\}\{y - \mathbb{E}[y]\}\right] \\ &= \mathbb{E}_{x,y}[xy] - \mathbb{E}[x]\mathbb{E}[y] \end{aligned}$$

$$\begin{aligned} \mathrm{cov}[\mathbf{x}, \mathbf{y}] &= \mathbb{E}_{\mathbf{x},\mathbf{y}}\left[\{\mathbf{x} - \mathbb{E}[\mathbf{x}]\}\{\mathbf{y}^{\mathrm{T}} - \mathbb{E}[\mathbf{y}^{\mathrm{T}}]\}\right] \\ &= \mathbb{E}_{\mathbf{x},\mathbf{y}}[\mathbf{x}\mathbf{y}^{\mathrm{T}}] - \mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{y}^{\mathrm{T}}] \end{aligned}$$

The Gaussian Distribution

$$\mathcal{N}\left(x|\mu,\sigma^2\right) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

x: μία μόνο πραγματική μεταβλητή
μ: μέσος , σ: τυπική απόκλιση



- Ικανοποιεί την ανισότητα

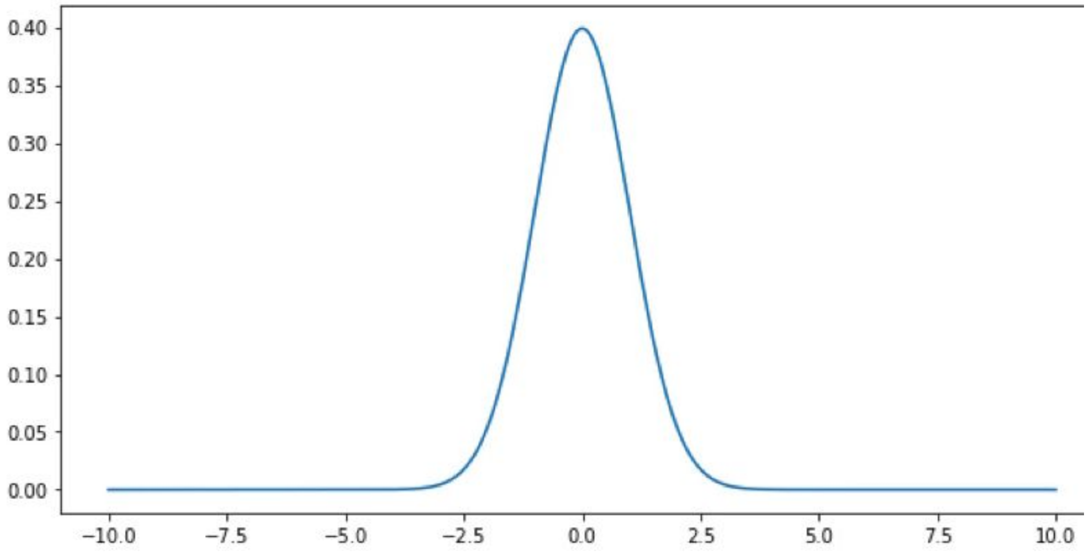
$$\mathcal{N}(x|\mu,\sigma^2) > 0$$

- Κανονικοποιημένη

$$\int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu,\sigma^2\right)\,\mathrm{d}x = 1$$

Flashback - Normal Distribution

Density $$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$



Gaussian Mean and Variance

Μέση τιμή του x ή μέσος :

$$\mathbb{E}[x] = \int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu,\sigma^2\right) x \,\mathrm{d}x = \mu$$

$$\mu = \mathbf{E}[x] \text{ hence } \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Διακύμανση : μέτρο μεταβλητότητας που υπάρχει στην

$$\mathrm{var}[x] = \mathbb{E}[x^2] - \mathbb{E}[x]^2 = \sigma^2$$

ρω από τη μέση τιμή της E[f(x)}

$$\sigma^2 = \mathbf{E}\left[(x-\mu)^2\right] \text{ hence } \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\mu})^2$$

Ροπή 2ης τάξης :

$$\mathbb{E}[x^2] = \int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu,\sigma^2\right) x^2 \,\mathrm{d}x = \mu^2 + \sigma^2$$

Συνάρτηση πιθανοφάνειας

Η συνάρτηση πιθανοφάνειας περιγράφει **πόσο πιθανό είναι να παρατηρήσουμε τα δεδομένα μας**, έχοντας ως δεδομένο **συγκεκριμένες παραμέτρους του μοντέλου**.

Συγκεκριμένα:

- Έστω ότι έχουμε ένα σετ παρατηρήσεων $x=(x1,x2,...,xn)$ και μια κατανομή πιθανότητας με παράμετρο θ.
- Η **συνάρτηση πιθανοφάνειας** $L(\theta)$ είναι η πιθανότητα να παρατηρήσουμε τα δεδομένα x, δεδομένης της παραμέτρου θ

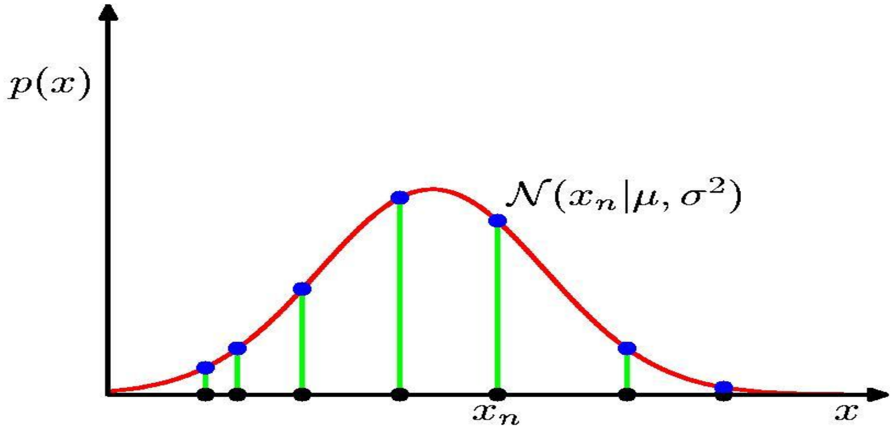
Για **διακριτές** τυχαίες μεταβλητές, η πιθανοφάνεια $L(\theta|x)$ ορίζεται ως: $L(\theta|x)=P(x|\theta)$

Για **συνεχείς** τυχαίες μεταβλητές, η συνάρτηση πυκνότητας πιθανότητας $f(x|\theta)$ χρησιμοποιείται αντί για την πιθανότητα P.

Πιθανότητα: τα δεδομένα είναι τυχαία και οι παράμετροι είναι σταθερές.

Πιθανοφάνεια: θεωρούμε τα δεδομένα σταθερά και ψάχνουμε για τις παραμέτρους που κάνουν αυτά τα δεδομένα πιο πιθανά

Gaussian Parameter Estimation



Likelihood function

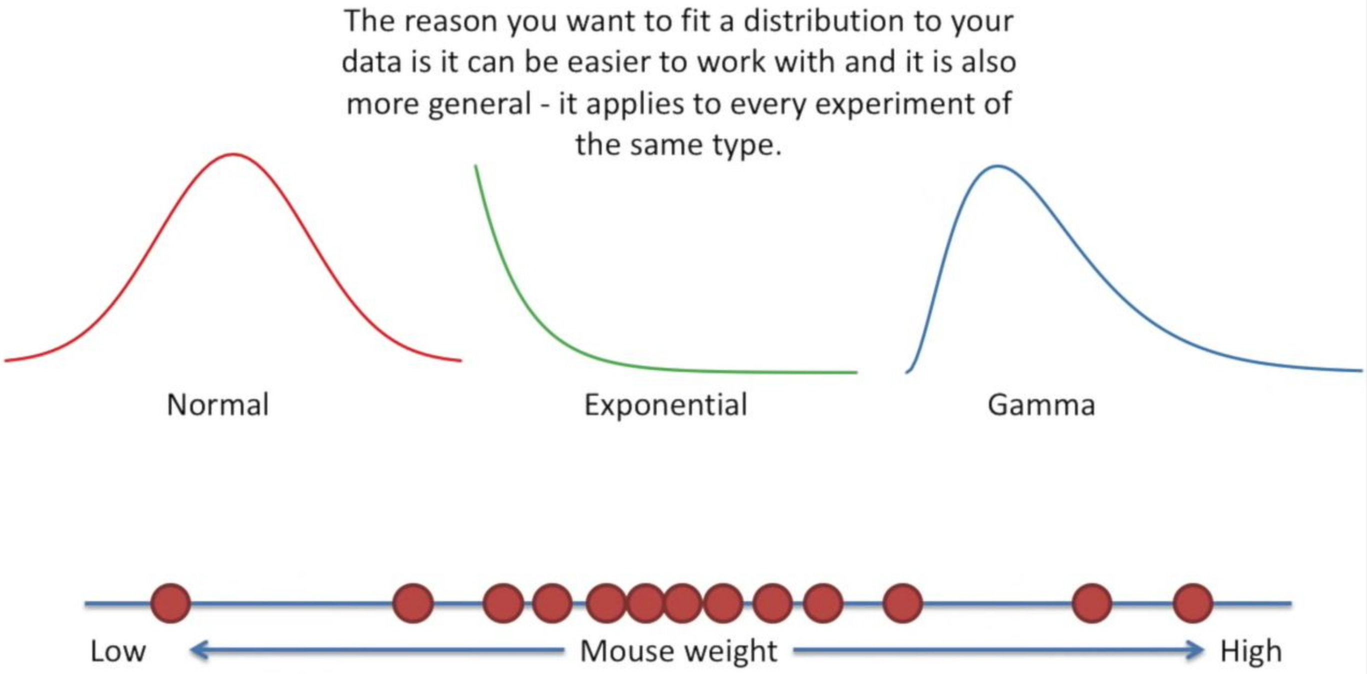
$$p(\mathbf{x}|\mu, \sigma^2) = \prod_{n=1}^{N} \mathcal{N}\left(x_n|\mu, \sigma^2\right)$$

$\mathbf{x}=(x_1,\ldots x_N)^T$

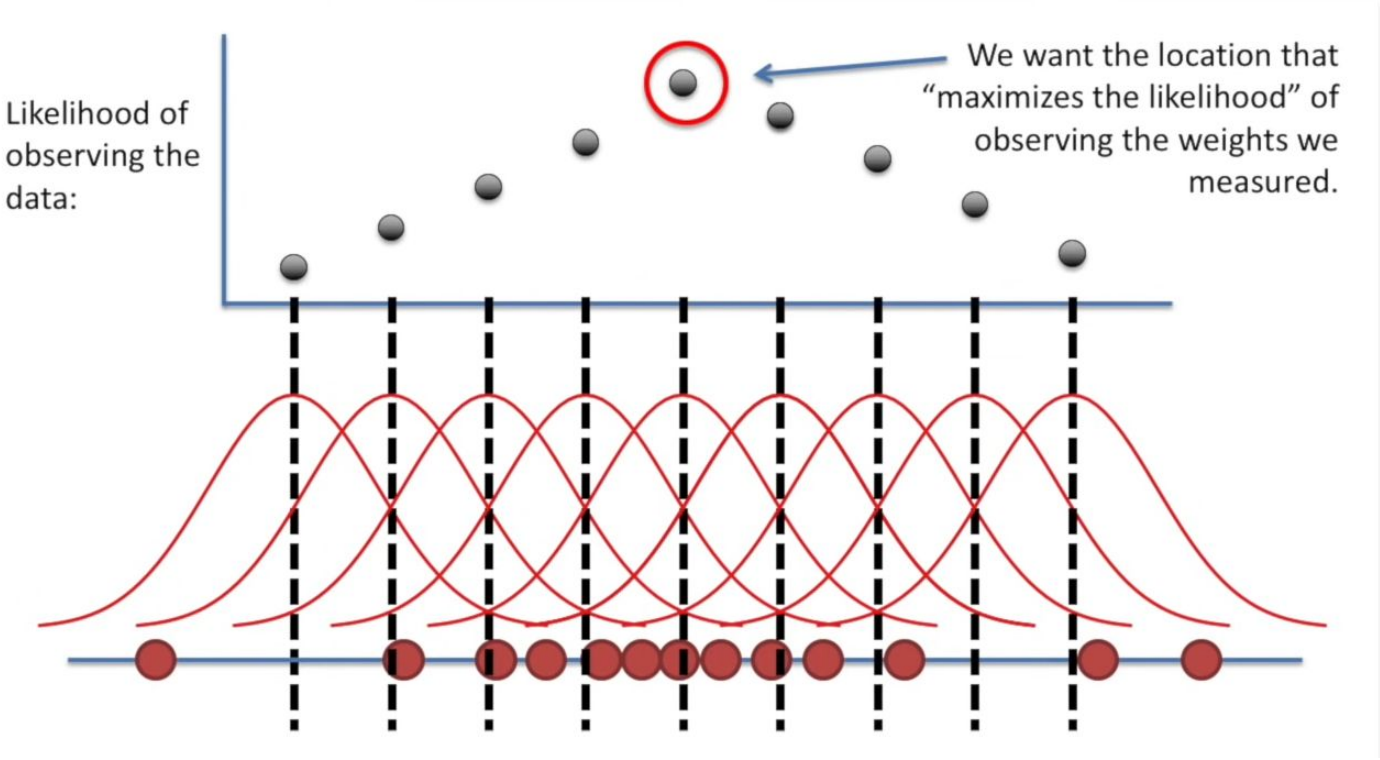
(x: i..i.d. Independent and identically distributed random variables)

Άγνωστα: μ,σ^2

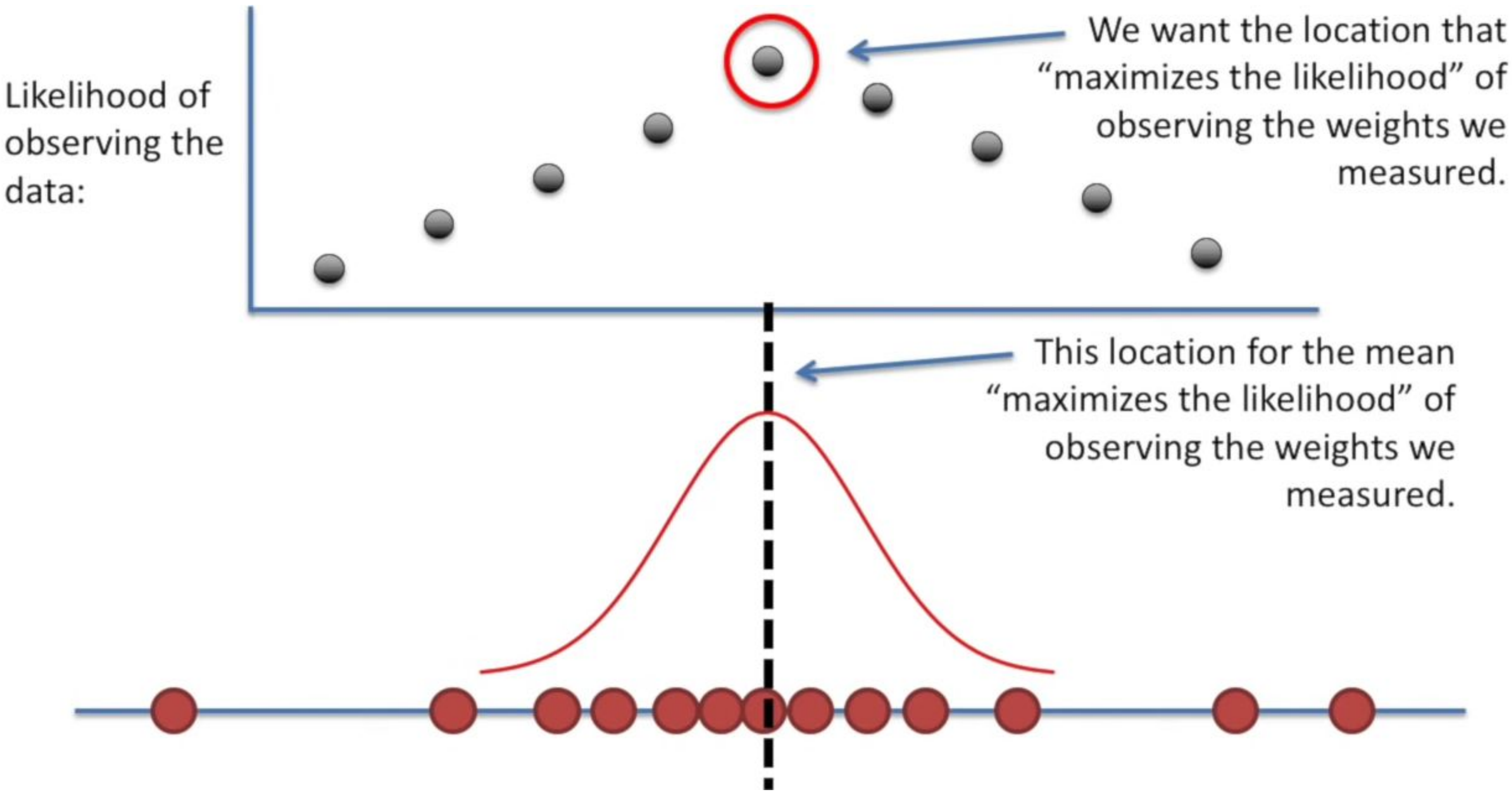
Maximum (Log) Likelihood



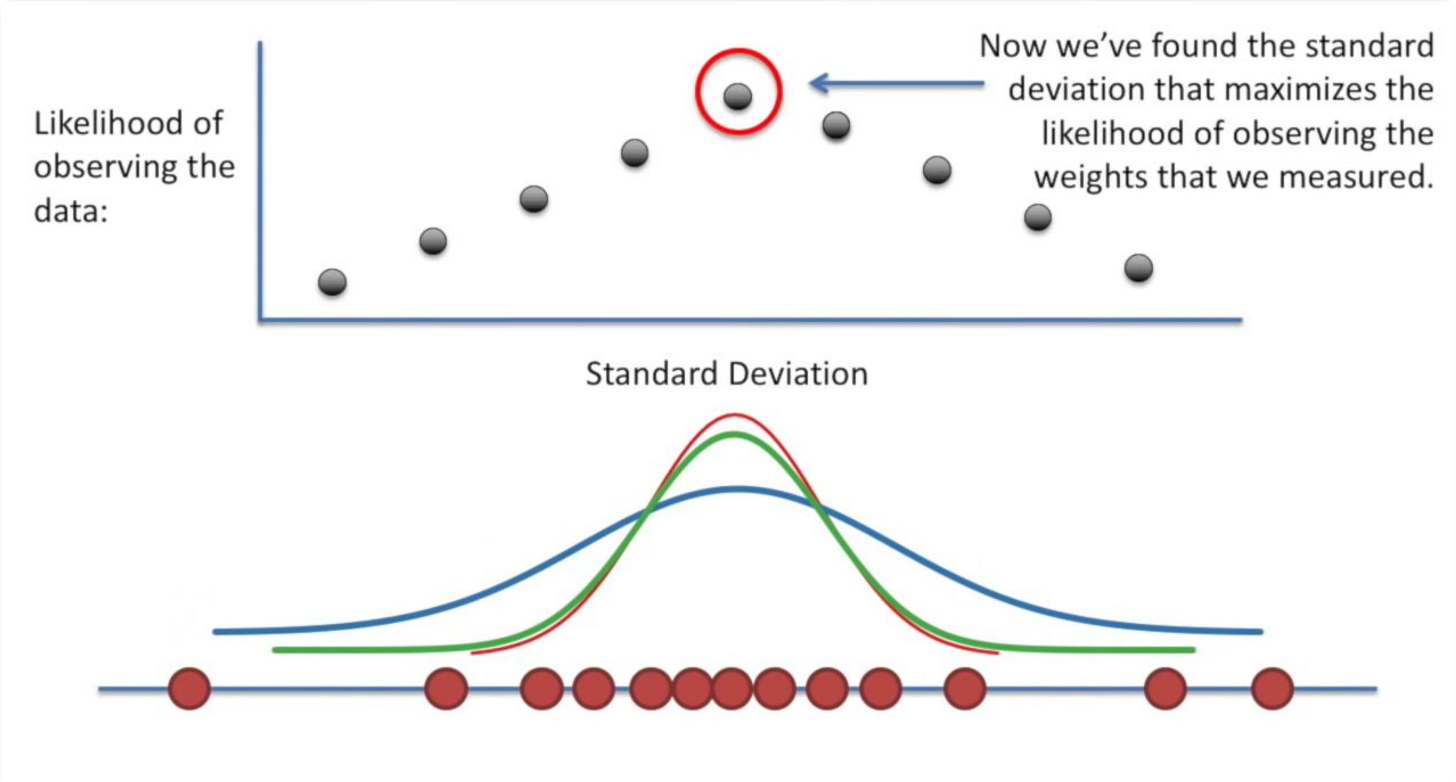
Maximum (Log) Likelihood



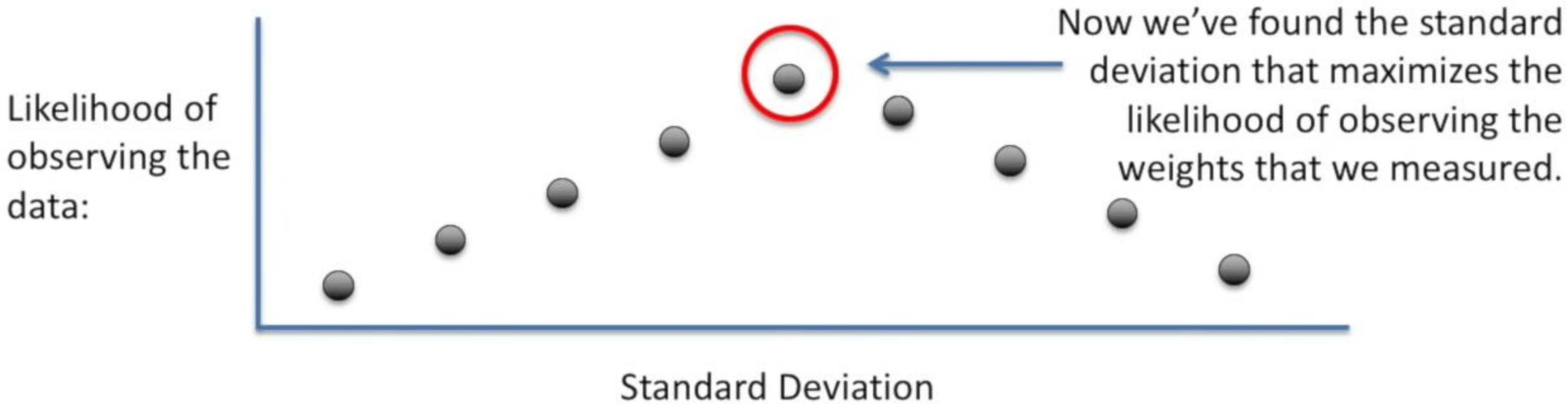
Maximum (Log) Likelihood

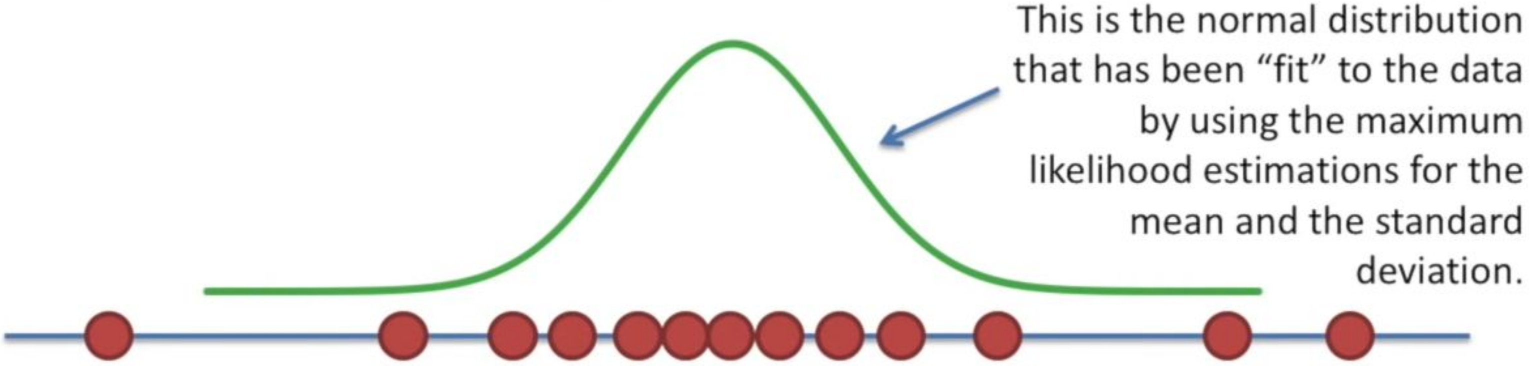


Maximum (Log) Likelihood



Maximum (Log) Likelihood





Likelihood

- Observe some data $X = \{x_1, \ldots x_n\}$
- Assume that the data is drawn from a Gaussian

$$p(X; \mu, \sigma^2) = \prod_{i=1}^{n} p(x_i; \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right)$$

- **Fitting parameters is maximizing** $p(X; \mu, \sigma^2)$ **wrt.** μ, σ^2
  (maximize likelihood that data was generated by model)
- **Practical simplification**

$$\underset{\mu,\sigma^2}{\text{maximize}}\, p(X; \mu, \sigma^2) \iff \underset{\mu,\sigma^2}{\text{minimize}} - \log p(X; \mu, \sigma^2)$$

Maximum Likelihood

- Estimate parameters by finding ones that explain the data

$$\underset{\mu,\sigma^2}{\text{minimize}} - \log p(X; \mu, \sigma^2)$$

- **Decompose likelihood**

$$-\log p(X; \mu, \sigma^2) = \sum_{i=1}^{n} \left(\frac{1}{2} \log(2\pi\sigma^2) + \frac{1}{2\sigma^2}(x_i - \mu)^2\right) = \frac{n}{2} \log(2\pi\sigma^2) + \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2$$

Minimize $\mu = \frac{1}{n} \sum_{i=1}^{n} x_i$

$$\boxed{p(X; \mu, \sigma^2) = \prod_{i=1}^{n} p(x_i; \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right)}$$

Maximum Likelihood

- Estimating the variance

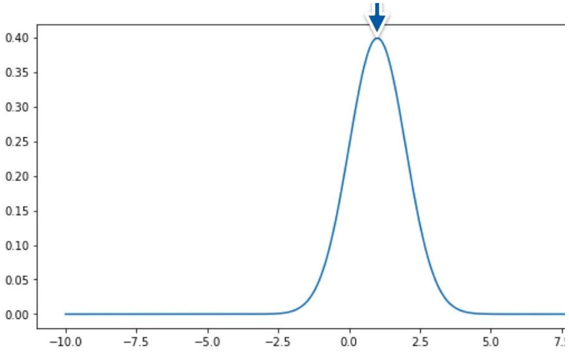
$$\frac{n}{2}\log(2\pi\sigma^2) + \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

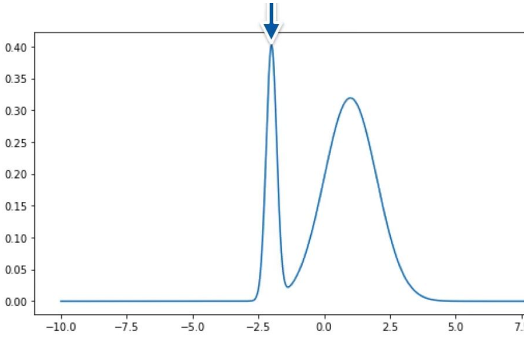
- Take derivatives with respect to it

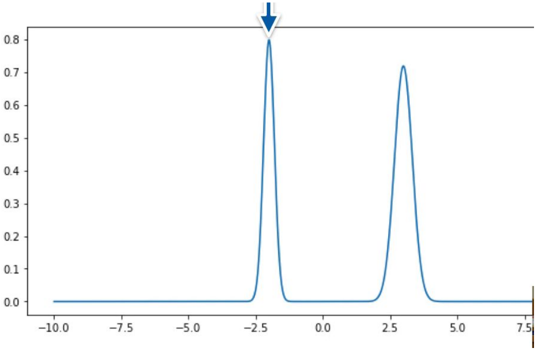
$$\partial_{\sigma^2}[\,\cdot\,] = \frac{n}{2\sigma^2} - \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2 = 0$$

$$\Longrightarrow \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$$

Maximum Likelihood Estimation







Κάτι μένει εκτός... χρησιμοποιώντας το MLE

Maximum Likelihood Estimation

- Data - ‘student didn’t do homework’
- Possible parameters
  - ‘dog ate homework’
  - ‘abducted by aliens’
  - ‘too lazy’
  - ‘sick grandmother’

- **All parameters explain the data.**

Maximize a posteriori Estimation

penalty

- Posterior Probability

$$p(w\,|\,X) \propto p(X\,|\,w)p(w)$$

hence $\; -\log p(w\,|\,X) = -\log p(X\,|\,w) - \log p(w) + c$

- Maximum a Posteriori Estimation

$$\underset{w}{\text{minimize}} -\log p(X;\mu,\sigma^2) - \log p(w)$$

- No homework example
p('no homework'|explanation) = 1 (all explanations work)

lazy student	grandma sick	dog ate it	alien abduct
0.8	0.19	0.0099	0.0001

Τι σχέση έχει το regression με το MLE;

- Recall optimization problem

$$\underset{w}{\text{minimize}} \sum_{i=1}^{n} (y_i - f(x_i, w))^2 + \text{penalty}(w)$$

$-\log p(w)$

Does the model work?

Additive Gaussian Noise

- Data generation model

$$y_i = f(x_i, w) + \epsilon_i \text{ where } \epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

- Gaussian Prior $p(w)$ hence $-\log p(w) = \frac{1}{2\bar{\sigma}^2} \|w\|^2$ +const

Regression

- Maximum a posteriori

$$\underset{w}{\text{minimize}} \; -\log p(w \mid X, Y)$$

$$\Longleftrightarrow \underset{w}{\text{minimize}} \; \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - f(x_i, w))^2 + \frac{1}{2\bar{\sigma}^2} \|w\|^2 + \text{const}$$

$$\Longleftrightarrow \underset{w}{\text{minimize}} \; \frac{1}{2n} \sum_{i=1}^{n} (y_i - f(x_i, w))^2 + \frac{\lambda}{2} \|w\|^2$$

Βιβλιογραφία

Strang Gilbert

→ Linear Algebra and Learning from Data (math.mit.edu/learningfromdata)

→ Introduction to Linear Algebra - Fifth Edition

→ MIT 18.065 Matrix Methods in Data Analysis, Signal Processing, and Machine Learning

Marc Peter Deisenroth, A. Aldo Faisal, and Cheng Soon Ong

→ **Mathematics for Machine Learning - Book**

→ **Mathematics for Machine Learning -github**

Ian Goodfellow, Yoshua Bengio, Aaron Courville

→ **Deep Learning,** Chapter 2: Linear Algebra

C. Bishop

→ Pattern Recognition and Machine Learning, , Chapter 1

2.3. Linear Algebra — Dive into Deep Learning 0.16.1 documentation