FROM LINEAR ALGEBRA TO MACHINE LEARNING

Omar Gutiérrez

July 27, 2018

@trinogz

Overview

Motivation

Vectors

Examples

Conclusions

MOTIVATION

Motivation

- **Linear algebra** is important to understand machine learning.
- As well as **calculus**, **probability theory**, and **statistics**.
- It is rewarding to take the **hard path** to learn machine learning (IMHO).

Learning from errors

```r
# compute distances using self-created function
distances2 <- matrix(nrow=size, ncol=size)
for (p in 1:size) {
  for (q in 1:size) {
    row_p = iris2[p,]
    row_q = iris2[q,]
    distances2[p, q] <- euclidean_distance(row_p, row_q)
  }
}
```

Learning from errors

```r
# function for calculating Euclidean Distance
euclidean_distance <- function(p, q) {
    ed = 0
    for (i in 1:4) {
        ed <- ed + (p[,i] - q[,i]) ^ 2
    }
    ed <- sqrt(ed)
    return(ed)
}
```

VECTORS

A vector is a collection of numbers

$$\vec{a} = \mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

Length of a vector

$$|\mathbf{a}| = \sqrt{\sum_{i=0}^{n} a_i^2}$$

Distance beetwen vectors

$$
\begin{aligned}
d(\boldsymbol{a}, \boldsymbol{b}) &= ||\boldsymbol{a} - \boldsymbol{b}|| \\
&= \sqrt{\sum_{i=0}^{n} (a_i - b_i)^2}
\end{aligned}
$$

Dot product

$$\mathbf{a} \cdot \mathbf{b} = \sum_{i=0}^{n} a_i b_i$$
$$= a_0 b_0 + a_1 b_1 + \ldots + a_n b_n$$

So,

$$\mathbf{a} \cdot \mathbf{a} = a_0 a_0 + a_1 a_1 + \ldots + a_n a_n$$
$$= a_0^2 + a_1^2 + \ldots + a_n^2$$
$$= |\mathbf{a}|^2$$

EXAMPLES

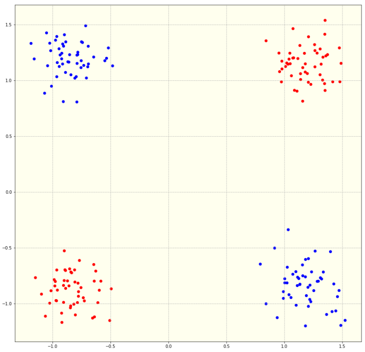
Winter is Coming

The AI winter is coming

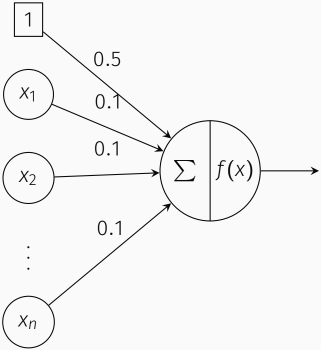
- Is really coming? No.
- However, we already had an AI winter.
- The research on neural nets was stopped for many years, after Minsky and Papert proved that a single layer perceptron was not able to deal with the exclusive-or problem.

Toy dataset

$$\begin{bmatrix} b_{11} & b_{12} & 1 \\ b_{21} & b_{22} & 0 \\ b_{31} & b_{32} & 1 \\ \dots & \dots & \dots \\ b_{n1} & b_{n2} & 1 \end{bmatrix}$$

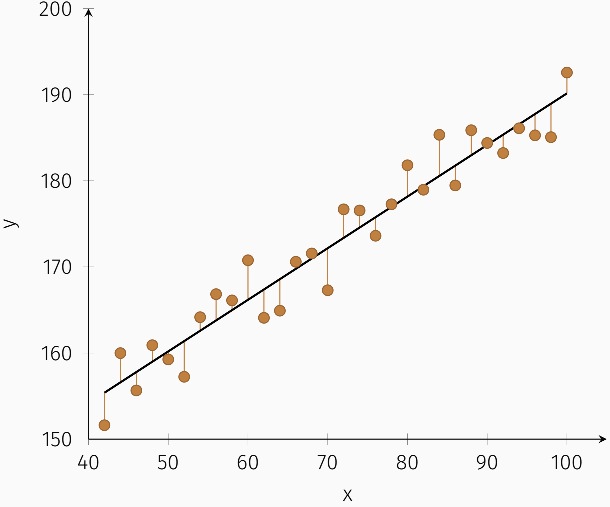


Perceptron



x_0	x_1	$\sum$	$f(x)$
1	1	$1 \times 0.5 + 1 \times -1 = -0.5$	0
1	0	$1 \times 0.5 + 0 \times -1 = 0.5$	1

Linear Regression



- We want to calculate the intercept a and the slope b.

$$\arg\min_{a,b} \sum_i (y_i - (ax_i + b))^2 = \arg\min_{\mathbf{w}} ||X\mathbf{w} - \mathbf{y}||^2$$

- The solution to this optimization problem is:

$$\mathbf{w}^* = (X^T X)^{-1} X^T \mathbf{y}\,.$$

CONCLUSIONS

More topics we should check

- Gradient descent is a beautiful optimization algorithm, basically, we multiply matrices to many times.
- Eigenvectors and eigenvalues; some dimensionality reduction techniques are based on eigendecomposition.
- Be aware that numerical instabilities can happen, and avoid these ones.

References

- Mathematics for Machine Learning: Linear Algebra by Coursera.
- The Math of Intelligence by Siraj Raval.
- Deep Learning Book by Bengio and Goodfellow, has a chapter summarizing which linear algebra topics you need to learn neural networks.

Thank you.
Questions?
Comments?