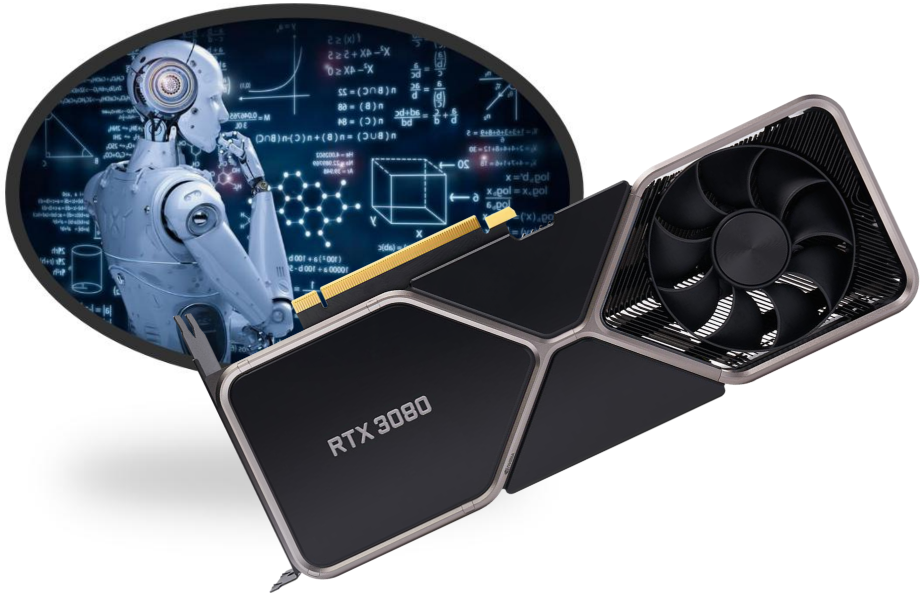

Machine Learning and Big Data Processing: Lab sessions

LAB4: PRESENTATION

Esther Rodrigo Bonet esther.rodrigo.bonet@vub.be (PL9.2.27)
Leandro Di Bella leandro.di.bella@vub.be (PL9.2.36)

Content

- **Sparse coding and dictionary learning**
- **Clustering**
  - K-means clustering
  - DBSCAN clustering

Sparse Coding and Dictionary Learning

Sparse Coding and Dictionary Learning

- Sparse representation is to find a sparse vector $\alpha \in R^m$ such that $x \approx D\alpha$

  $\alpha \in R^m$: sparse code

  D : a set of normalized "basis vectors" ($D = [d_1, d_2, ..., d_m] \in R^{n \times m}$)

  $x \in R^n$: a signal

- The Sparse Coding Model
  - The optimization problem

$$\tilde{a} = \arg\min_{a} \underbrace{\frac{1}{2}\|x - Da\|_2^2}_{\text{Data fitting term}} + \lambda \underbrace{g(a)}_{\text{Regularization term}}$$

– *The L_2 norm* $\|a\|_2^2 = \sum_{i=1}^{m} a_i^2$

– *The L_0 norm* $\|a\|_0 = |\mathbb{S}|, \quad \mathbb{S} = \{a_i | a_i \neq 0\}$

– *The L_1 norm* $\|a\|_1 = \sum_{i=1}^{m} |a_i|$

Sparse Coding and Dictionary Learning

- Orthogonal Matching Pursuit
  - The optimization problem is:

$$\tilde{a} = \arg\min_{a} \underbrace{\frac{1}{2}\|x - Da\|_2^2}_{\textit{Data fitting term}} \quad \text{s.t.} \quad \underbrace{\|a\|_0 \le L}_{\textit{Regularization term}}$$

Sparse Coding and Dictionary Learning

- Orthogonal Matching Pursuit
  - The optimization problem is:

$$\tilde{a} = \arg\min_{a} \underbrace{\frac{1}{2}\|x - Da\|_2^2}_{\textit{Data fitting term}} \quad \text{s.t.} \quad \underbrace{\|a\|_0 \leq L}_{\textit{Regularization term}}$$

- Steps for solving
  - Initialization: $a = 0$ residual $r = x$ active set $\Omega = \phi$
  - while $\|a\|_0 < L$
  - Select the element with maximum correlation with the residual
  - Update the active set, coefficients and residual
  - end while

Sparse Coding and Dictionary Learning

- Dictionary Learning
  - The optimization problem is:

$$\underset{D,A}{\text{minimize}} \quad \underbrace{\frac{1}{2}\left\|X - DA\right\|_F^2}_{\textit{Data fitting term}}, \quad \text{s.t.} \quad \underbrace{\left\|a_t\right\|_0 \leq s}_{\textit{Regularization term}}, \quad \forall t = 1, 2, \ldots T.$$

Sparse Coding and Dictionary Learning

- Dictionary Learning
  - The optimization problem is:

$$\underset{D,A}{\text{minimize}} \quad \underbrace{\frac{1}{2}\left\|X - DA\right\|_F^2}_{\text{Data fitting term}}, \quad \text{s.t.} \quad \underbrace{\left\|a_t\right\|_0 \leq s}_{\text{Regularization term}}, \quad \forall t = 1, 2, \ldots T.$$

- Solutions
  - Given initial estimates for dictionary D
  - iterate on k between the updates: Sparse coding step and Dictionary update step
    - ***Sparse coding step***

$$A^{k+1} = \arg\min_{A} \frac{1}{2}\|X - DA\|_F^2, \ \text{s.t.} \ \|a_t\|_0 \leq s_t \quad \forall t = 1, 2, \ldots T.$$

    - ***Dictionary update***

$$D^{k+1} = \arg\min_{D} \ \frac{1}{2}\|X - DA\|_F^2 \longrightarrow \boxed{D^{k+1} = XA^{k+1^T}\left(A^{k+1}A^{k+1^T}\right)^{-1}}$$

Clustering

K-means clustering

- Introduction
  - K-means is most commonly used clustering algorithm based on Euclidean distance.

$$D_{ij} = d(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}) = \|\boldsymbol{x}^{(i)} - \boldsymbol{x}^{(j)}\|_2^2 = \sum_{k=1}^{m} \left(x_k^{(i)} - x_k^{(j)} \right)^2$$

- The k-means algorithm
  1. Start with an initial set of centroids (choose K data samples at random from the input dataset)
  2. Assign each data sample to the closest centroid
  3. Re-compute the centroid of each cluster
  4. Repeat steps 2 and 3 until convergency
- Try multiple random initializations and run K-means multiple times

DBSCAN clustering

- DBSCAN : Density-Based Spatial Clustering of Applications with Noise.
- The DBSCAN algorithm

```
Input: The data set D

Parameter: ε, MinPts

For each object p in D
    if p is a core object and not processed then
        C = retrieve all objects density-reachable from p
        mark all objects in C as processed
         report C as a cluster
    else mark p as outlier
    end if

End For
```