Parallel Computing on Depression Analysis

Presentator: Yanan Mao

Content

- Introduction
- Methodology
  - Transfer Learning
  - Parallel K-means clustering Algorithm
- Results
- Discussion
- References

1

Introduction

Depression

- Depression is not the blues, or sadness or simply down, it is also a lasting overwhelming negative.
- Depression causes feelings of sadness and/or a loss of interest in activities you once enjoyed. It can lead to a variety of emotional and physical problems and can decrease your ability of function at work and at home.

What is depression: https://www.youtube.com/watch?v=z-IR48Mb3W0

"I'm Fine" - Learning To Live With Depression: https://www.youtube.com/watch?v=IDPDEKtd2yM

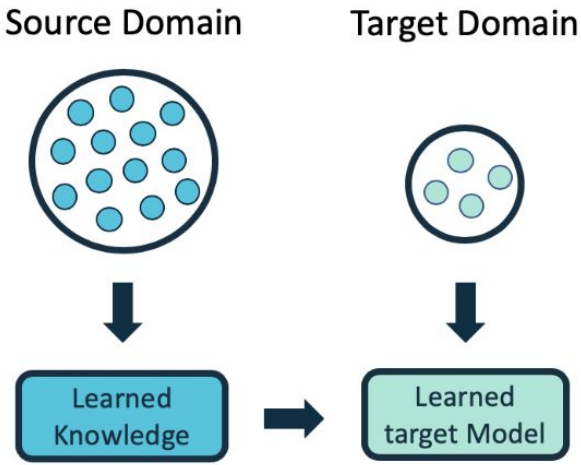
How to cope with Depression: https://www.youtube.com/watch?v=8Su5VtKeXU8

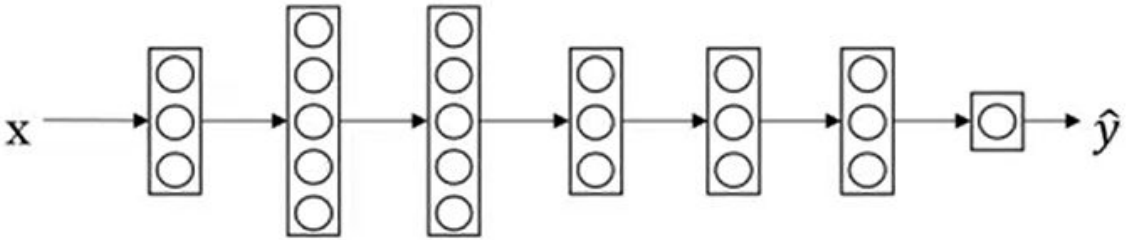
2

Methodology

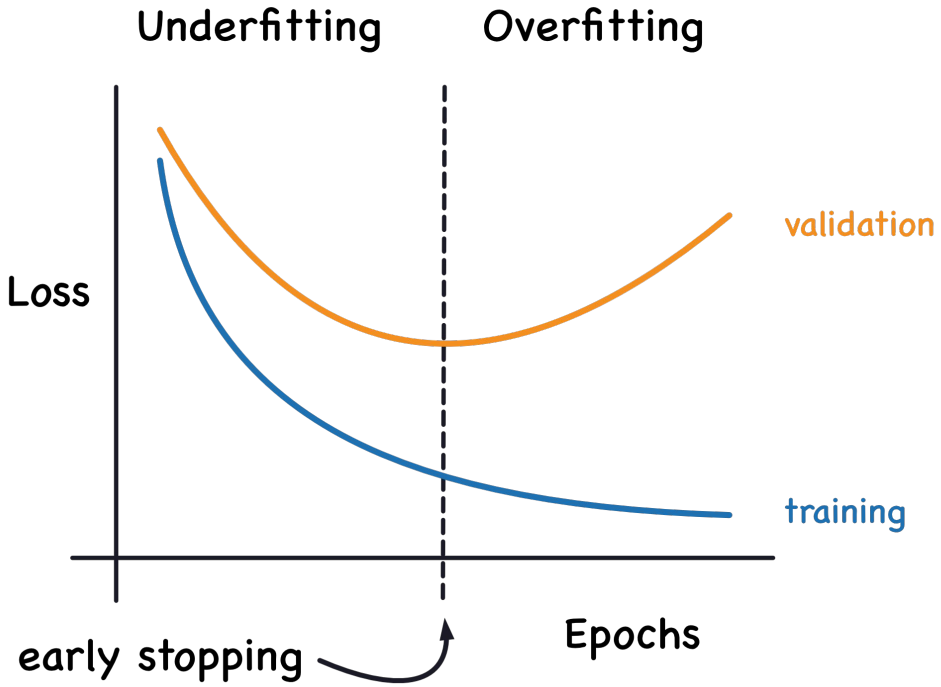
Transfer Learning

Take the knowledge that neural network have learnt from source domain, and apply that knowledge to target domain.



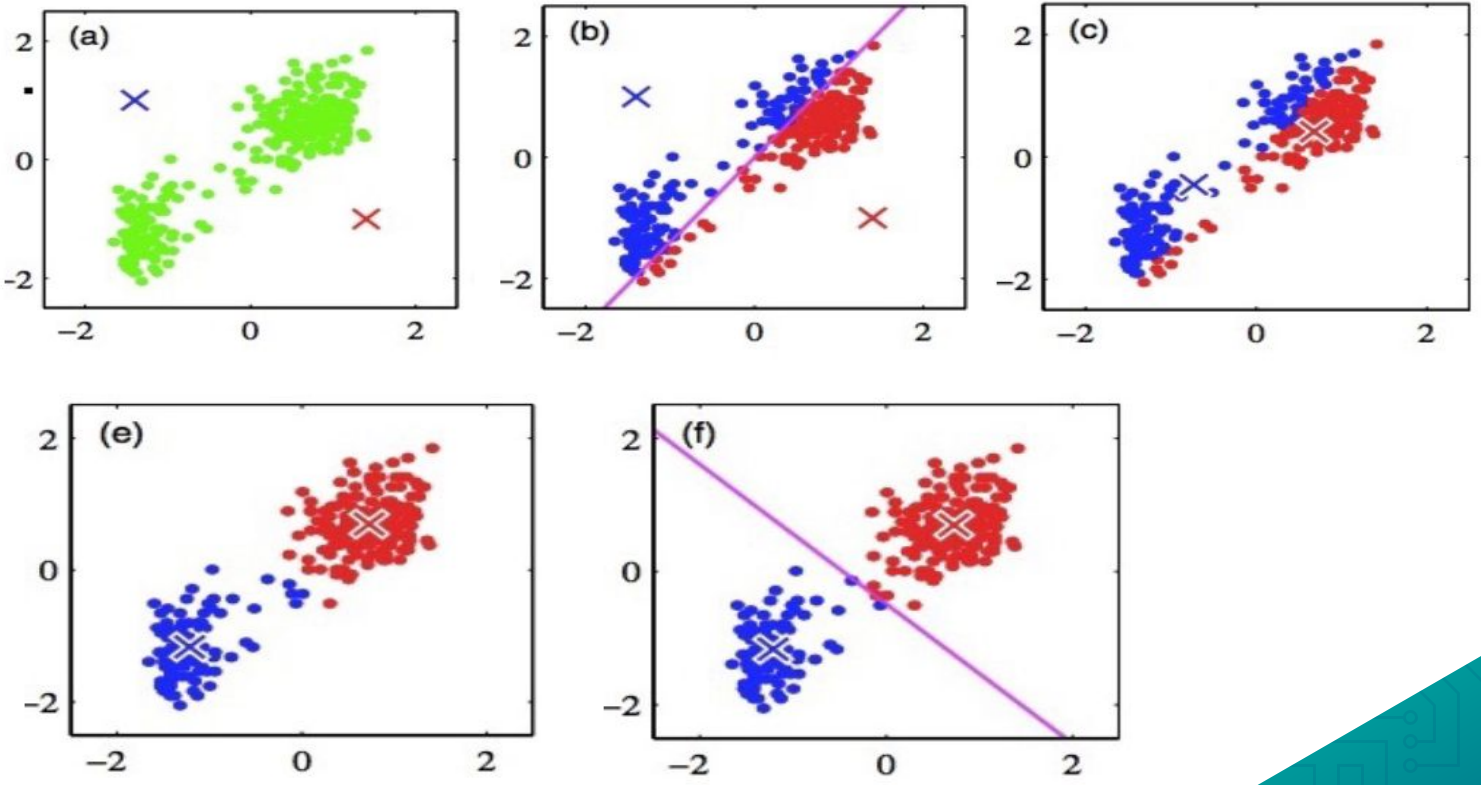


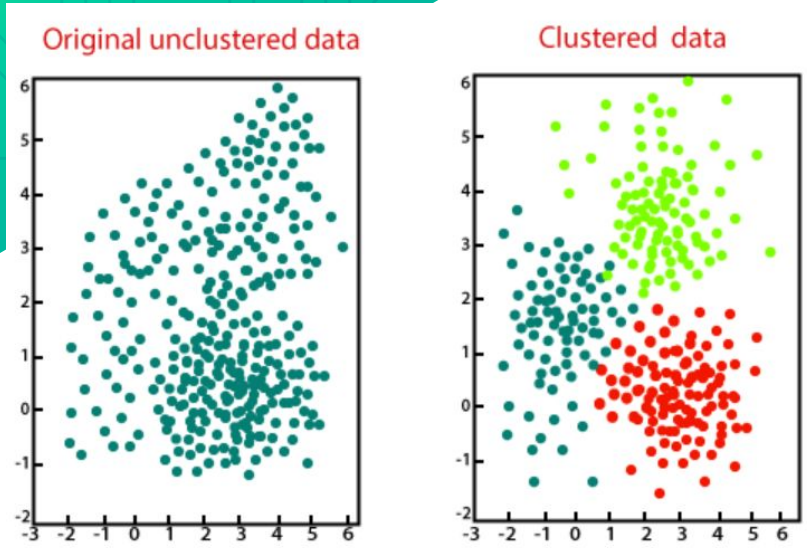
Underfit and Overfit

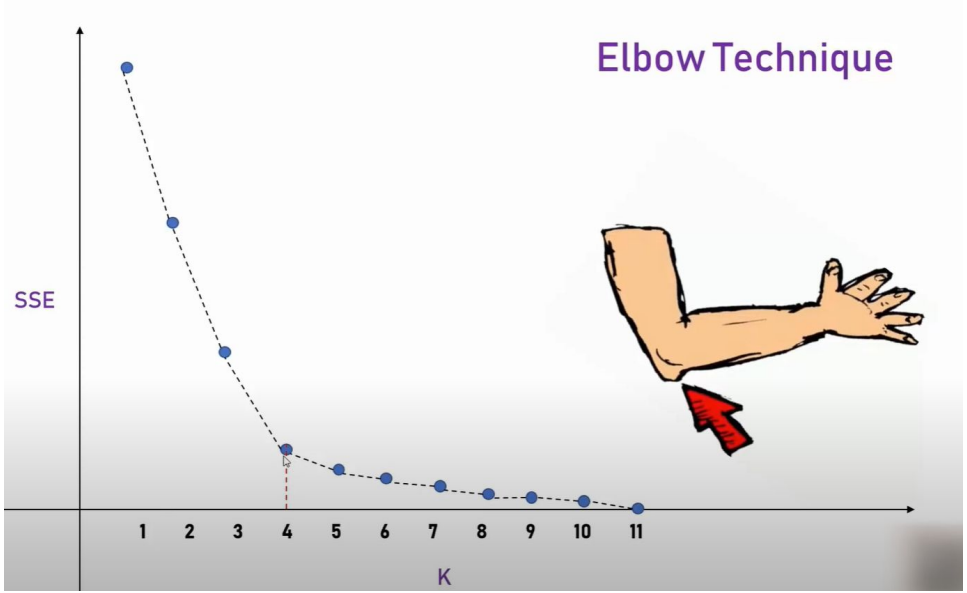


K-means

Example: K=2



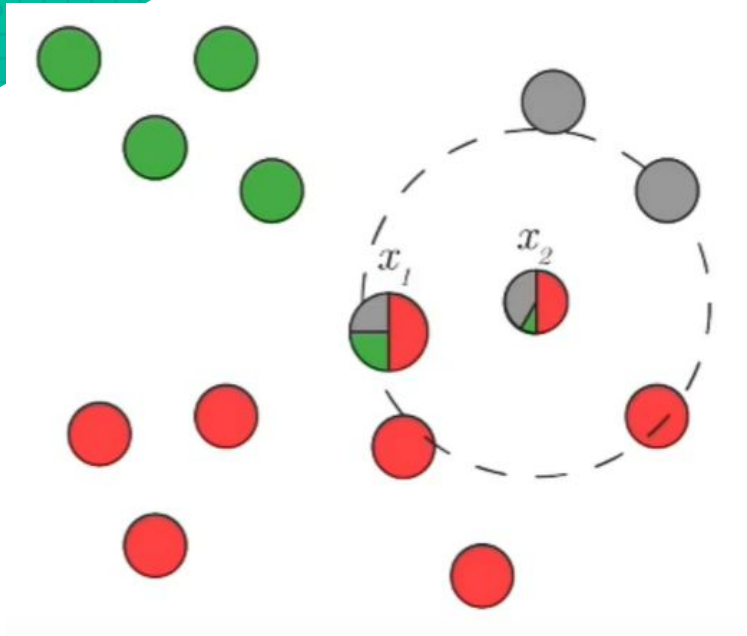




Related links: https://www.youtube.com/watch?v=hDmNF9JG3lo

https://www.youtube.com/watch?v=4b5d3muPQmA

Parallel k-means



Algorithm 1: K-means‖

Input: Dataset $\mathbf{X}$, #clusters K, and over-sampling factor l.
Output: Set of prototypes $\mathbf{C} = \{\mathbf{c}_1, \mathbf{c}_2, ..., \mathbf{c}_K\}$.

1: $\mathbf{C} \leftarrow$ select point $\mathbf{c}_1$ uniformly random from $\mathbf{X}$.
2: $\psi \leftarrow$ compute $SSE(\mathbf{C})$.
3: **for** $O(\log(\psi))$ times **do**
4: $\quad \mathbf{C}' \leftarrow$ sample each point $\mathbf{x} \in \mathbf{X}$ independently with probability $l \cdot d(\mathbf{x})^2 / SSE(\mathbf{C})$.
5: $\quad \mathbf{C} \leftarrow \mathbf{C} \cup \mathbf{C}'$
6: For each $\mathbf{x}$ in $\mathbf{C}$ attach a weight defined as the number of points in $\mathbf{X}$ closer to $\mathbf{x}$ than any other point in $\mathbf{C}$.
7: Do a weighted clustering of $\mathbf{C}$ into K clusters.

3

Results

Accuracy Table

Source domain	Heart	Breast Cancer	Lung Cancer
Train Accuracy	81.76%	?	?
Test Accuracy	79.04%	?	?
# of Samples	32	?	?

4

Discussion

- Any question you have?
- The relationship between source domain and target domain.
- Distinguish overfitting and underfitting?
- How to find the elbow point?And why we need that?
- What are the differences between k-means and parallel k-means?
- What will you do when you feeling down. Are you feeling not well recently?

5

References

- J. Hämäläinen, T. Kärkkäinen and T. Rossi, "Scalable initialization methods for large-scale clustering", *arXiv:2007.11937*, 2020
- Olman V, Mao F, Wu H, Xu Y: Parallel Clustering Algorithm for Large Data Sets with applications in Bioinformatics. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 2007, in press.
- Amy Kim, Parallel k-means from scratch. 2020
- Timothy G. Dinan John F. Cryan Gilliard Lach, Harriet Schellekens. Anxiety, depression, and the microbiome: A role for gut peptides. Neurotherapeutics, 15:36–59, 2018.