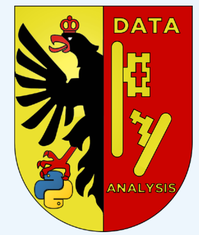


MEETUP #4 - SGD

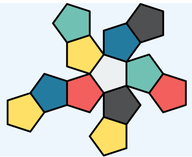
VISUALISING HIGHER DIMENSIONAL DATA USING **T-SNE**

Yacine Haddad

20/11/2018

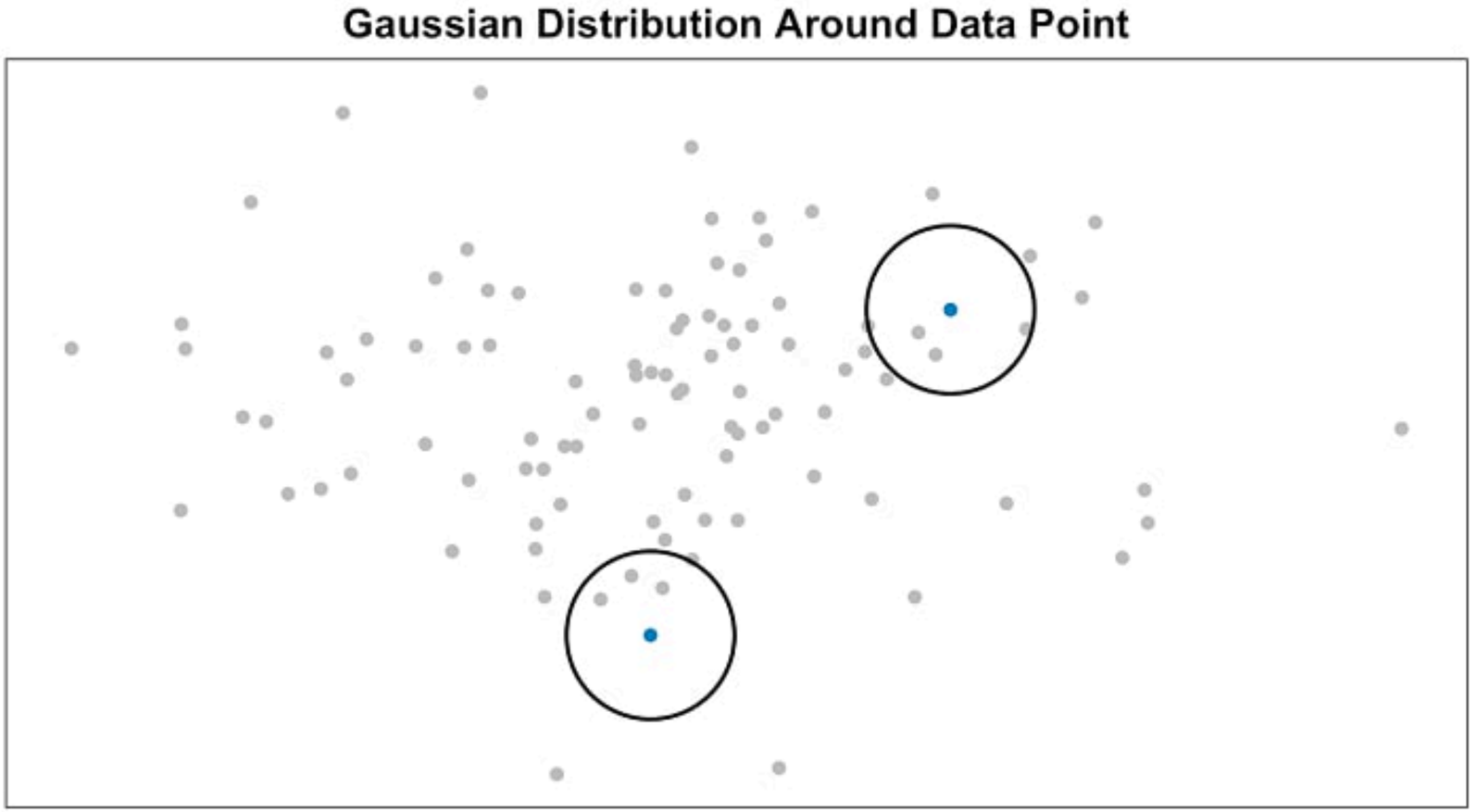
- **What is t-SNE?**

t-Distributed Stochastic Neighbour Embedding (t-SNE) is an unsupervised, non-linear techniqueprimarily used for data exploration and visualising high-dimensional data. In simpler terms, t-SNE gives you a feel or intuition of how the data is arranged in a high-dimensional space. It was developed by Laurens van der Maatens and Geoffrey Hinton in 2008.



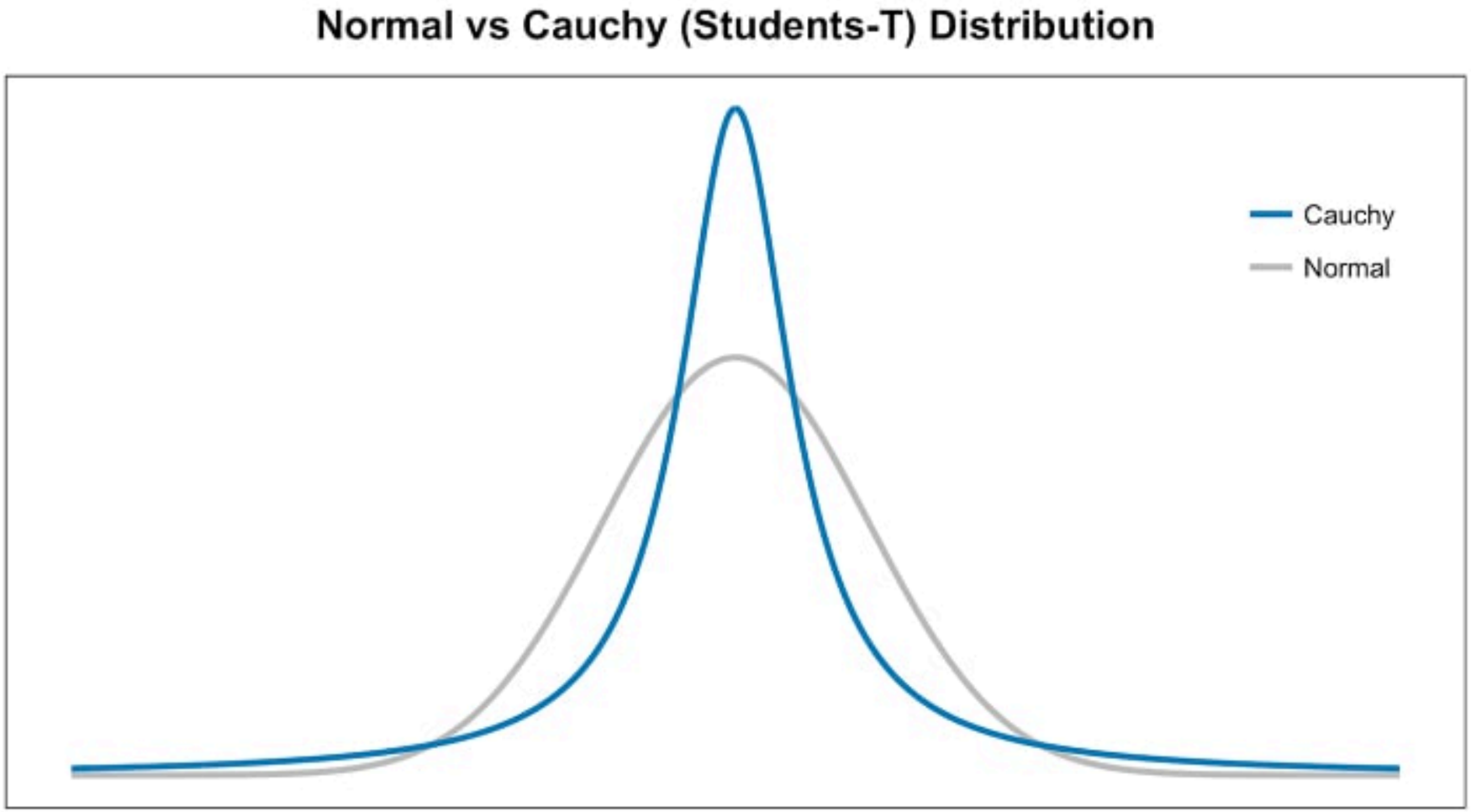
STEP 1

- Measure similarities between points in the high dimensional space
- Centre a Gaussian distribution over that point
- Measure the density of all points under the Gaussian distribution



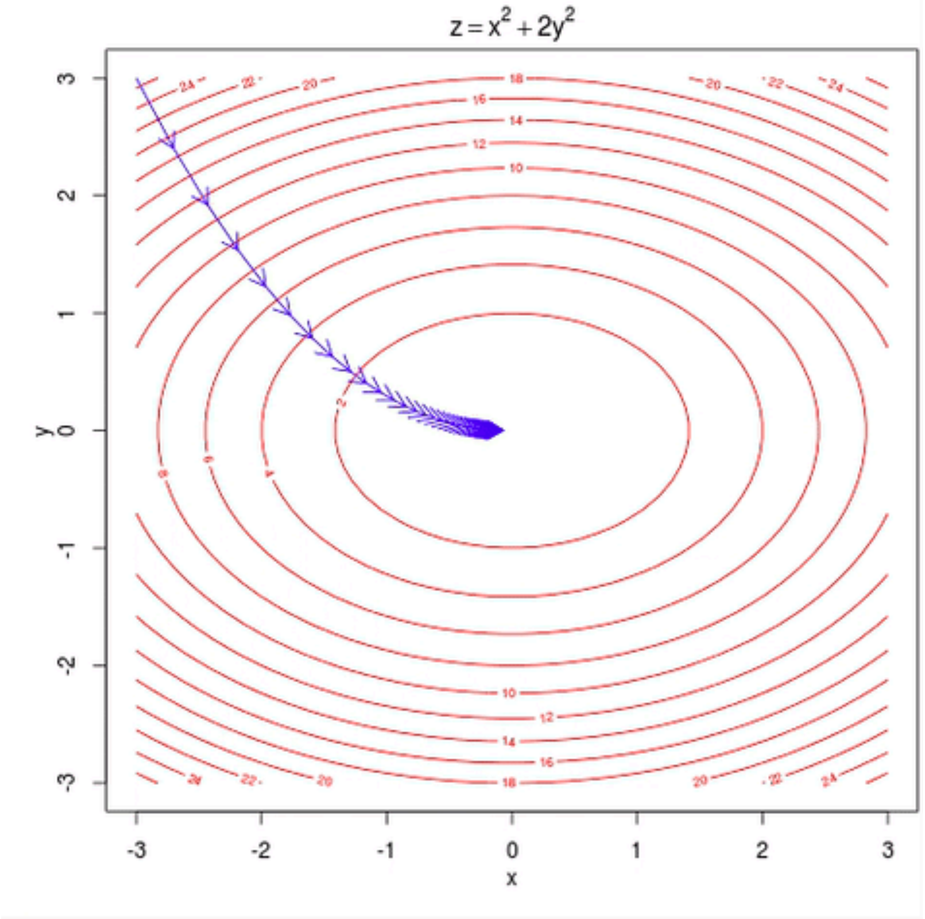
STEP 2

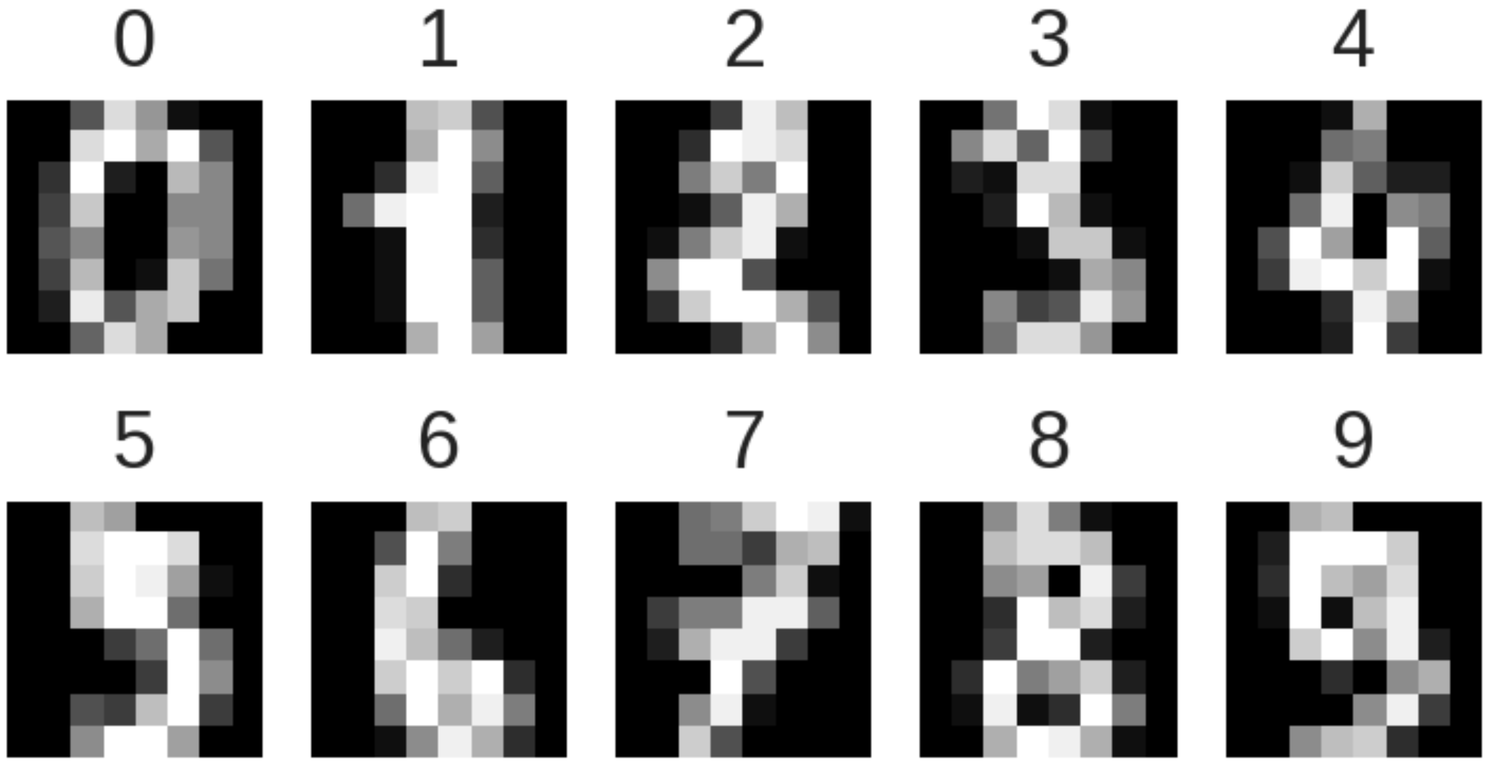
- **Instead of using a Gaussian, use t-distribution**

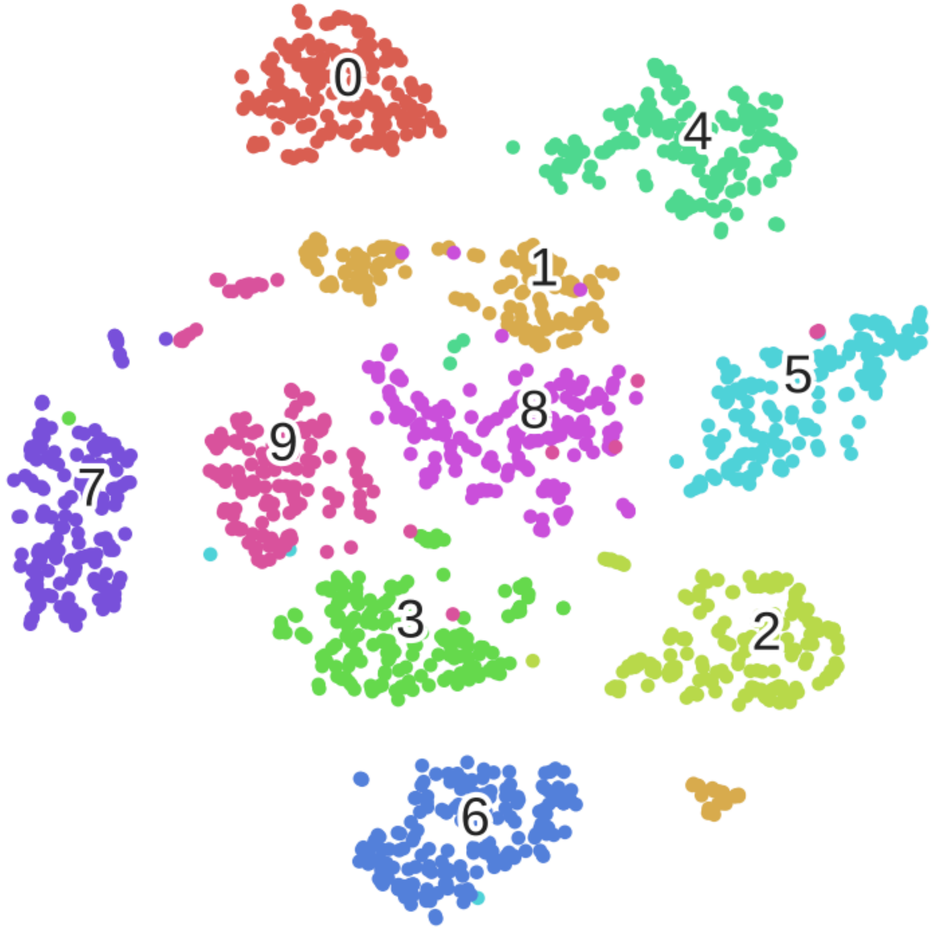


STEP 3

- Combine the set of probabilities in a Kullback-Liebler divergence cost function
- Use gradient descent to minimise cost function




0
1
2
3
4
5
6
7
8
9


0
4
1
5
8
9
7
3
2
6

- TSNE example : https://github.com/SocieteGenevoiseDonnees/MultivariateVisualisation/blob/master/tSNE_introduction.ipynb
- Google's Tensorflow embedding tool: https://projector.tensorflow.org/
- https://en.wikipedia.org/wiki/T-distributed_stochastic_neighbor_embedding
- https://www.kdnuggets.com/2018/08/introduction-t-sne-python.html
- https://scikit-learn.org/stable/auto_examples/manifold/plot_t_sne_perplexity.html#sphx-glr-auto-examples-manifold-plot-t-sne-perplexity-py
- https://indico.io/blog/visualizing-with-t-sne/
- Very detailed : https://www.analyticsvidhya.com/blog/2017/01/t-sne-implementation-r-python/
- Author's page : https://lvdmaaten.github.io/tsne/
- R package: https://www.rdocumentation.org/packages/tsne/versions/0.1-3/topics/tsne