Advanced Course on Deep Generative Models

Lecture 3: Latent Variable Models
Gaussian Mixture Models (GMM)
Expectation Maximization (EM)

Dan Rosenbaum, CS Haifa

Today

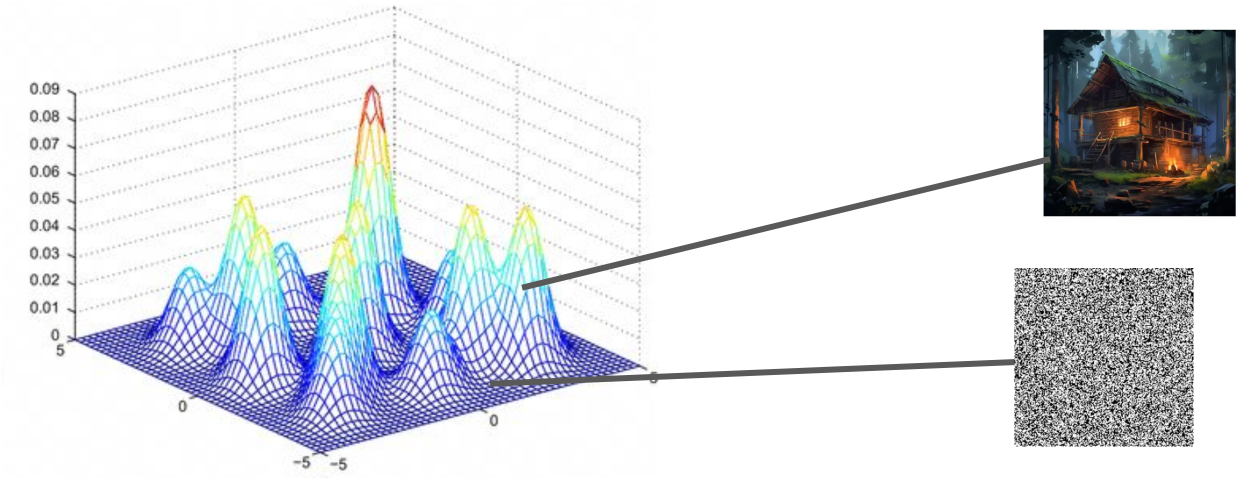
- Recap – Bayesian Inference
- latent variable models
- Gaussian Mixture Models (GMM)
- Expectation Maximization (EM)

What can we do with generative models?

- Solve some task (e.g. generative classifier)
- Generate data
- Representation learning
- Make decisions
- Probabilistic inference

Learning a generative model

- Data is generated by an unknown underlying distribution $\mathbf{p_{data}}$
- We are looking for the parameters $\boldsymbol{\theta}$ such that $\mathbf{p}_{\boldsymbol{\theta}}$ is close to $\mathbf{p_{data}}$



Learning a generative model

- Find $\boldsymbol{\theta}$ with Maximum likelihood - classical (frequentist) approach
    - Equivalent to minimizing KL divergence between $\mathbf{p}_{\mathbf{data}}$ and $\mathbf{p}_{\boldsymbol{\theta}}$
- Find posterior over $\boldsymbol{\theta}$ - Bayesian approach
- Using latent variables is somewhere in the middle.

The Bayesian Philosophy

- Every unknown value is a random variable
- We always maintain a distribution over random variable ⇒ “belief”
- Given some observation, the distribution is updated via Bayes’ rule

Advantages: capture uncertainty, can define optimal estimators.

Disadvantages: assumes a prior, can be hard to compute.

Linear Gaussian models

Consider the problem where $\mathbf{y} = \mathbf{A}\boldsymbol{\theta} + \boldsymbol{\eta}$, where $\boldsymbol{\eta}$ is a random vector with a Gaussian distribution $\mathcal{N}(0, \mathbf{I}\,\boldsymbol{\sigma}^2)$.

Maximizing the likelihood is a general form of:

1. Linear regression with squared loss (the rows of **A** are the values of **x**)

2. Polynomial regression (the rows of **A** consist of $\mathbf{x}, \mathbf{x}^2, \mathbf{x}^2, \ldots, \mathbf{x}^k$)

Solving for $\boldsymbol{\theta}$:

Linear Gaussian models

$$
\begin{aligned}
y \sim & \mathcal{N}(A\theta, \sigma^2 I) \\
\frac{\partial \log p_\theta(y)}{\partial \theta} = & \frac{1}{\sigma^2} A^\top (y - A\theta) \\
\frac{\partial \log p_\theta(y)}{\partial \theta} = & 0 \\
A^\top = & A^\top A\theta \\
\hat{\theta}_{\mathrm{M}L} = & (A^\top A)^{-1} A^\top y
\end{aligned}
$$

Bayesian Linear Gaussian Models

Bayesian Inference

Gaussian models for images

Why not use a Gaussian model for images?

1. For large images we will need a huge covariance matrix

2. Distribution of images is clearly non-Gaussian

How can we Construct Models of Images?

1. Conditional independence of pixels

   → autoregressive models: next pixel prediction

2. Conditional independence using additional variables

   → latent variable models

3. Constrained parameterization

   → continuous representation
   → function approximation (e.g. with neural networks)

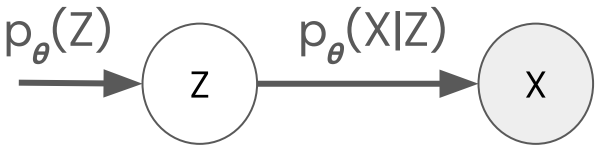
Bayes Nets

Conditional independence on pixels

Can be effective, but not very natural structure.

More Semantic Structure

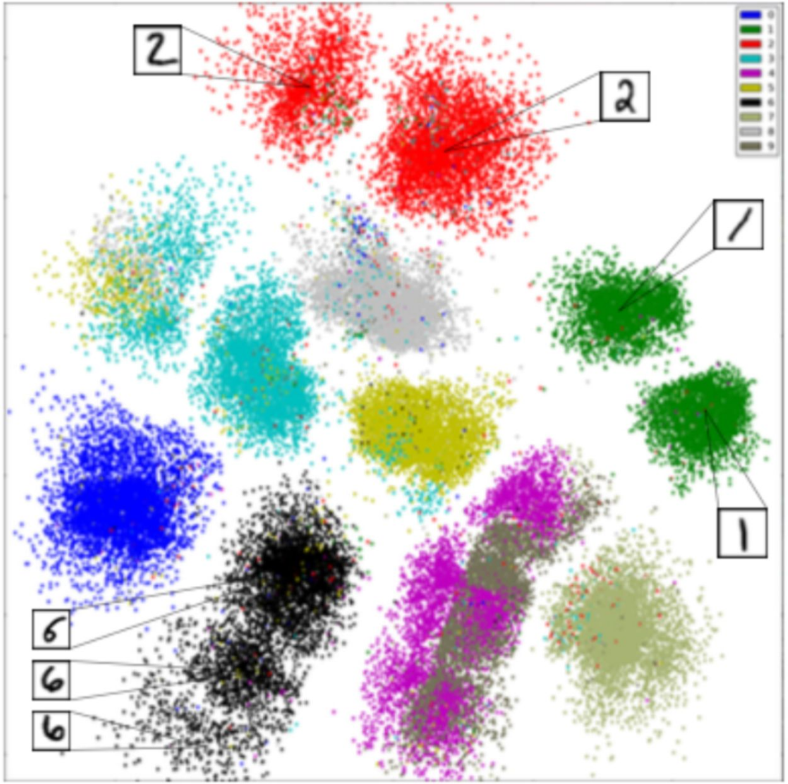
Latent Variable Models



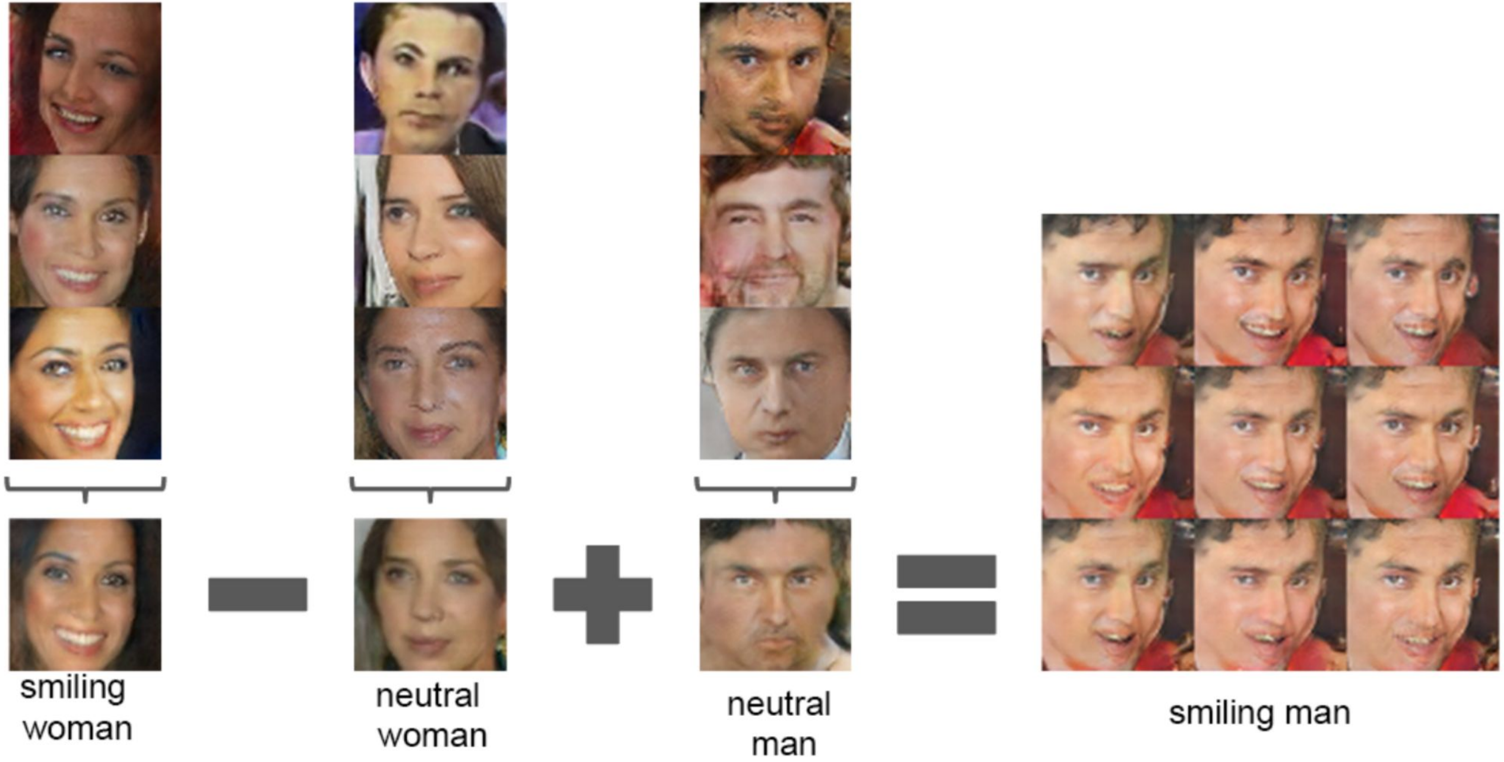
- Assume there’s an additional variable which we don’t see (latent, hidden)
- Use it to construct an efficient conditional independence structure
- This can solve the issues we had (efficient non-Gaussian models)
- Has a more natural interpretation
- Semi-Bayesian approach:
  - Unknown latent variable **Z** + unknown parameters $\boldsymbol{\theta}$.
  - Maximum likelihood for $\boldsymbol{\theta}$, Bayesian for **Z**

Latent Variable Model

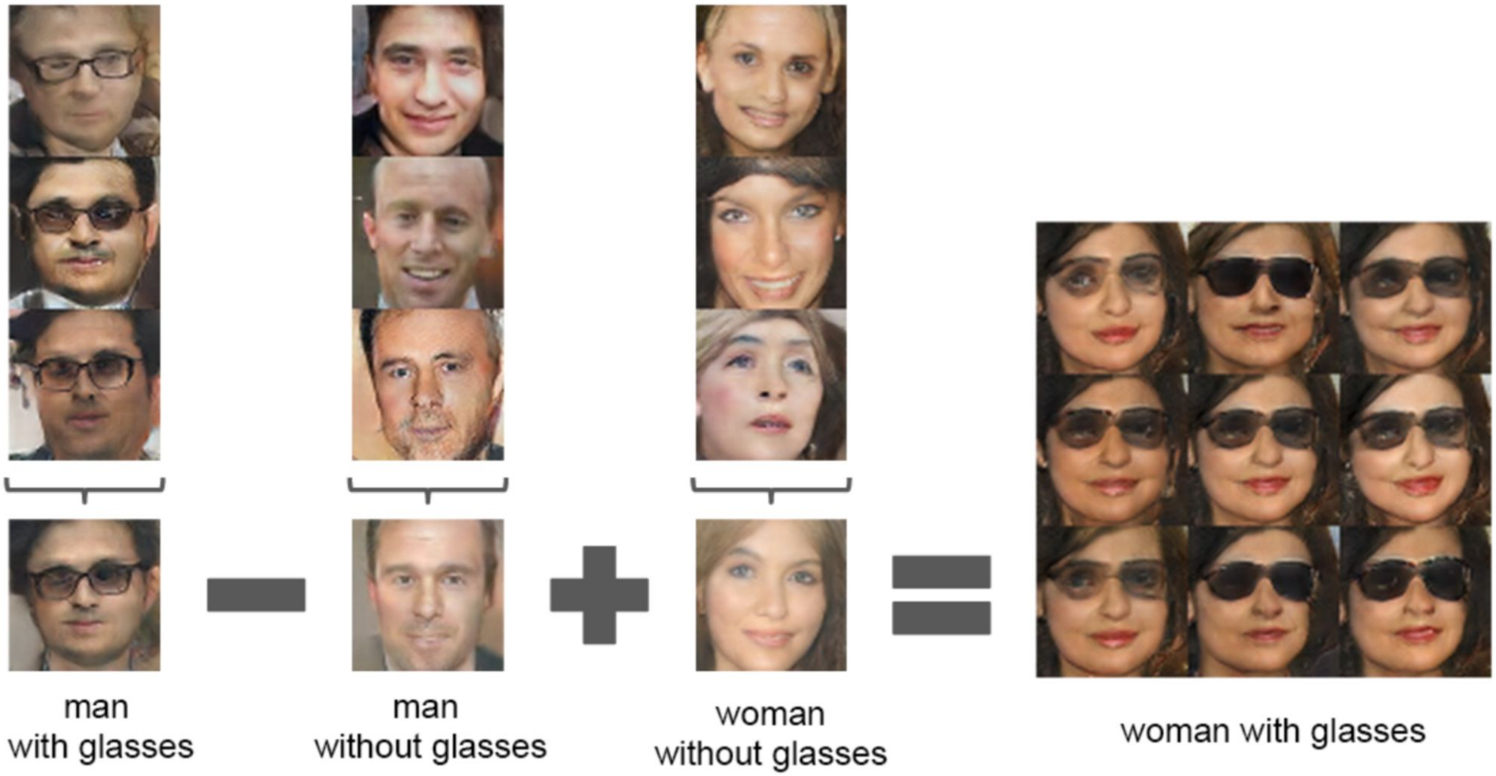
Representation Learning with the Latent Vector



Representation Learning with the Latent Vector

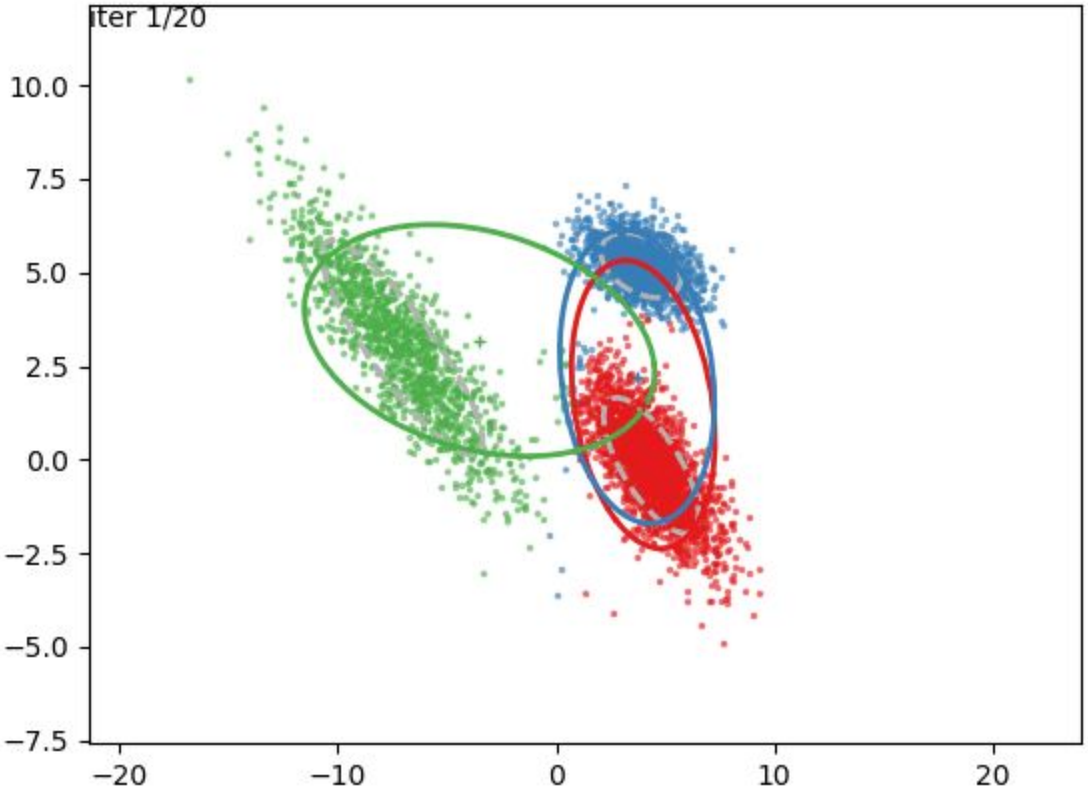


Representation Learning with the Latent Vector

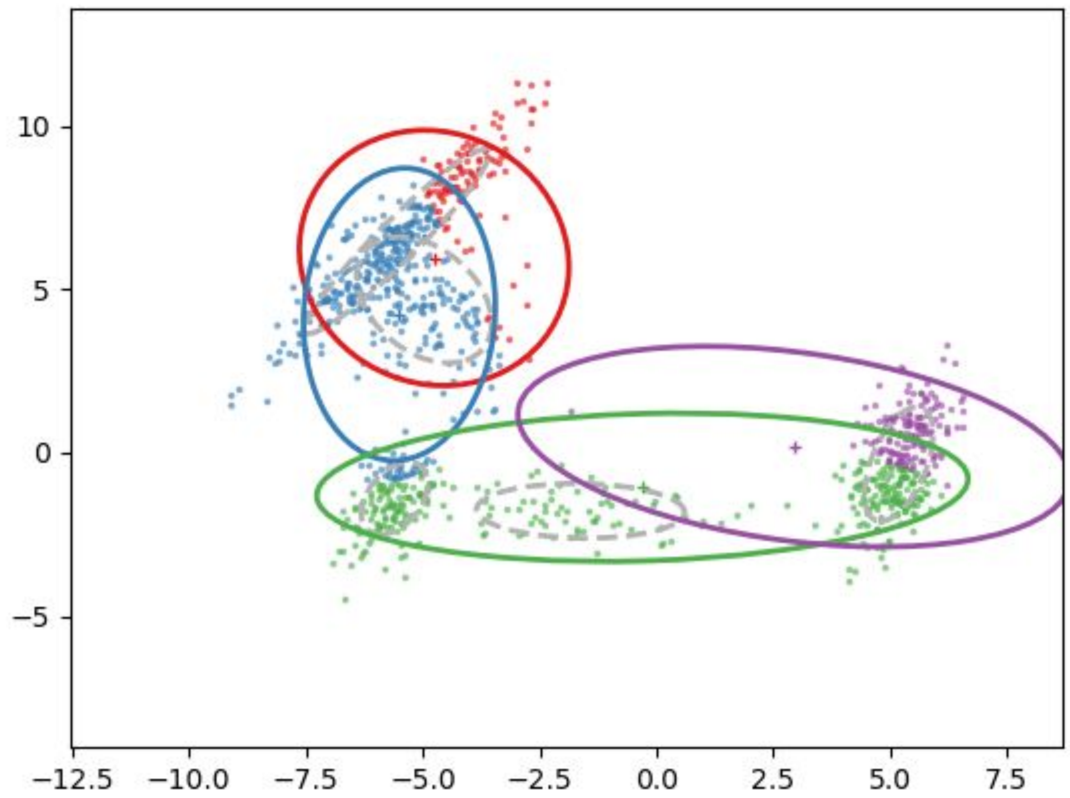


Directly on images

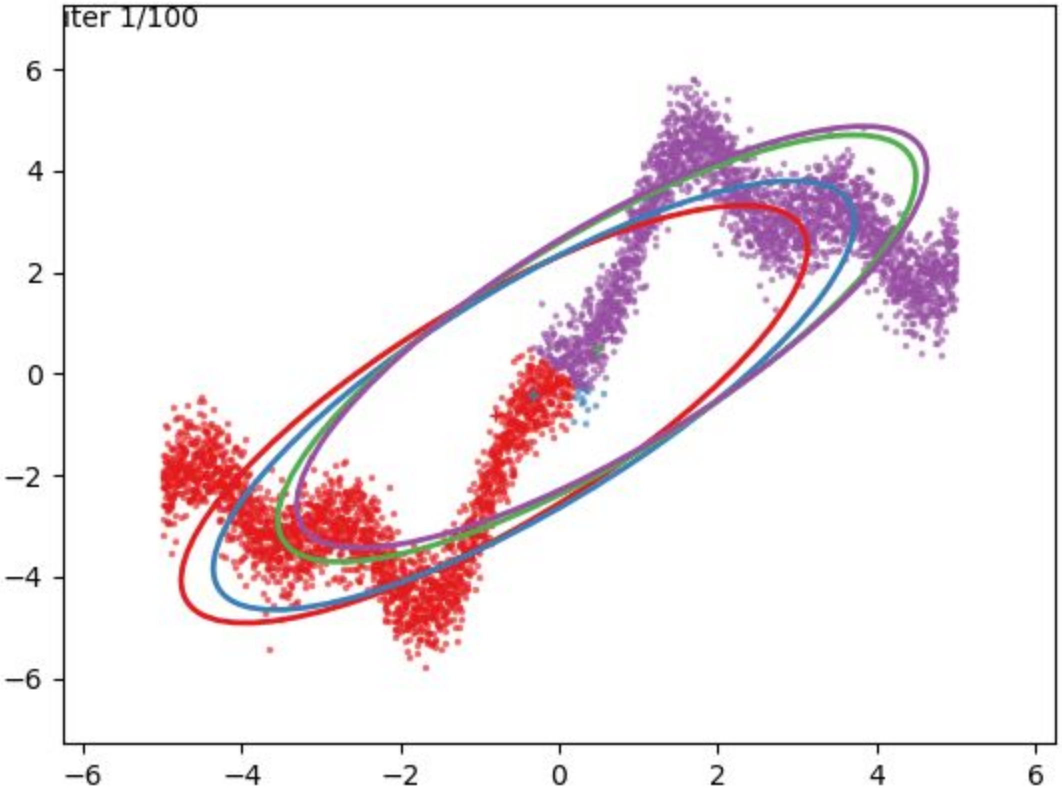
Results of doing the same arithmetic in pixel space



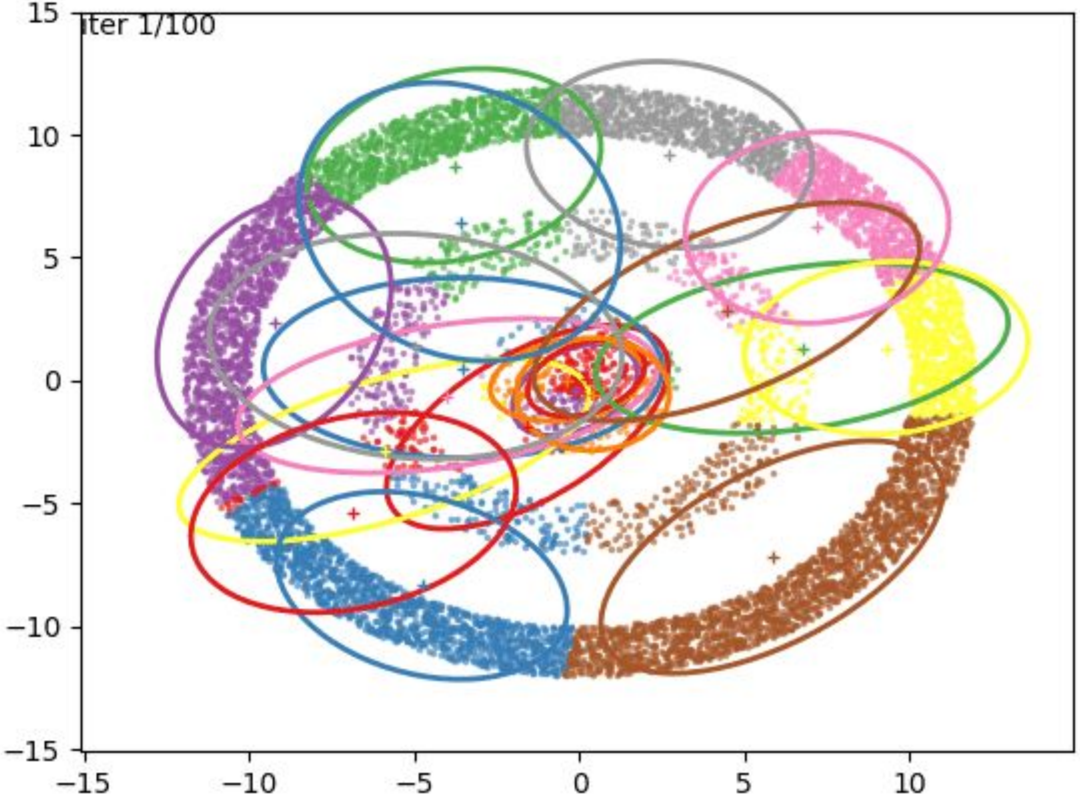
Credit: Roy Friedman



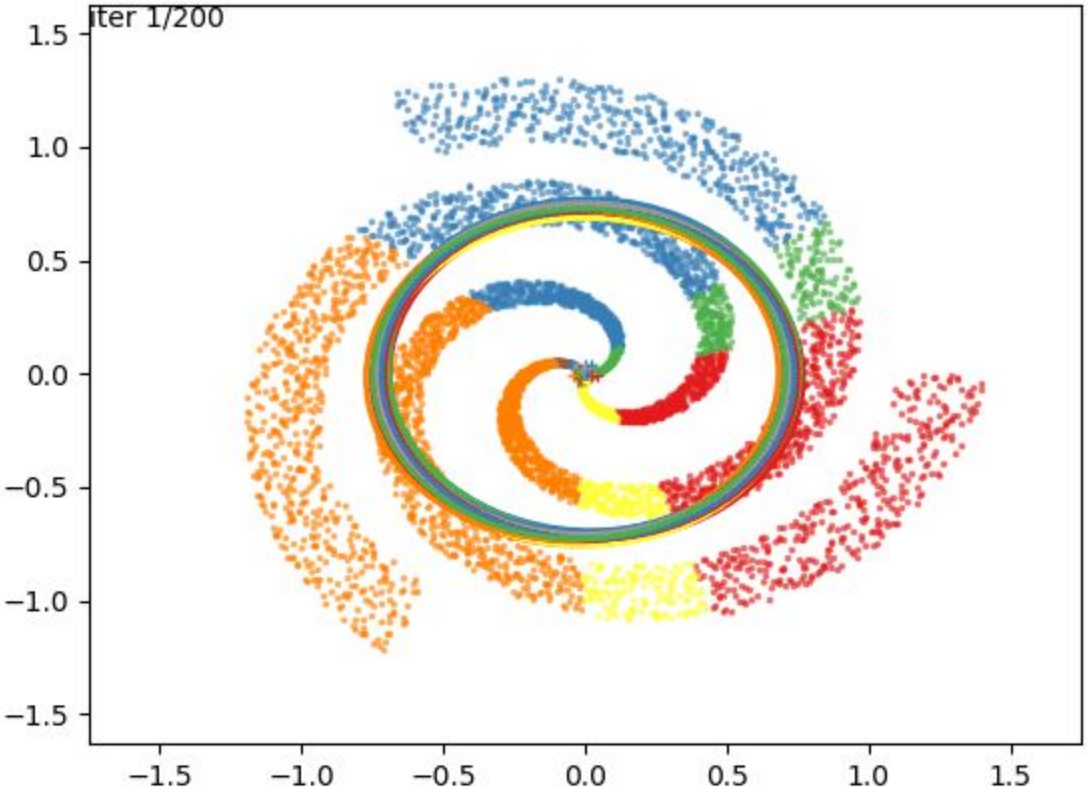
Credit: Roy Friedman



Credit: Roy Friedman



Credit: Roy Friedman



Credit: Roy Friedman