

Progress Report

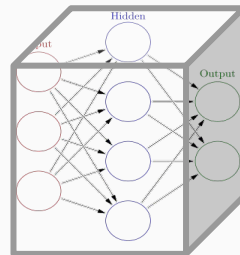
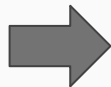
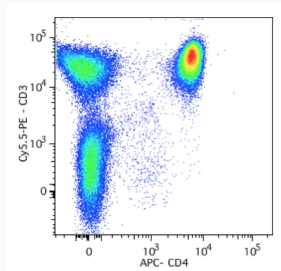
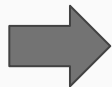
Kuei-Yueh Ko



Goal

Group 01
(ex: Control)

Group 02
(ex: Disease)



Group 01
or
Group 02
?

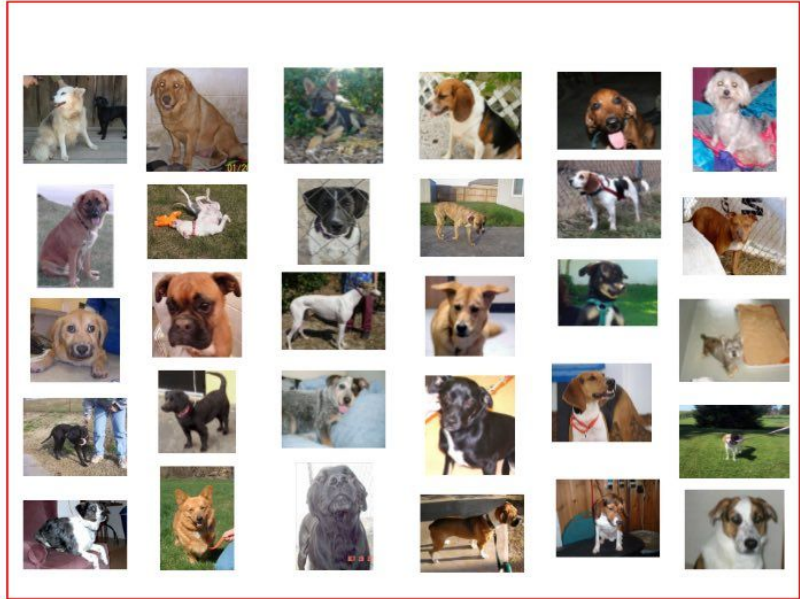
In short, it is a supervised learning problem

Image -> Convolutional Neural Network (CNN)

Cats



Dogs



Sample of cats & dogs images from Kaggle Dataset

Image -> Convolutional Neural Network (CNN)

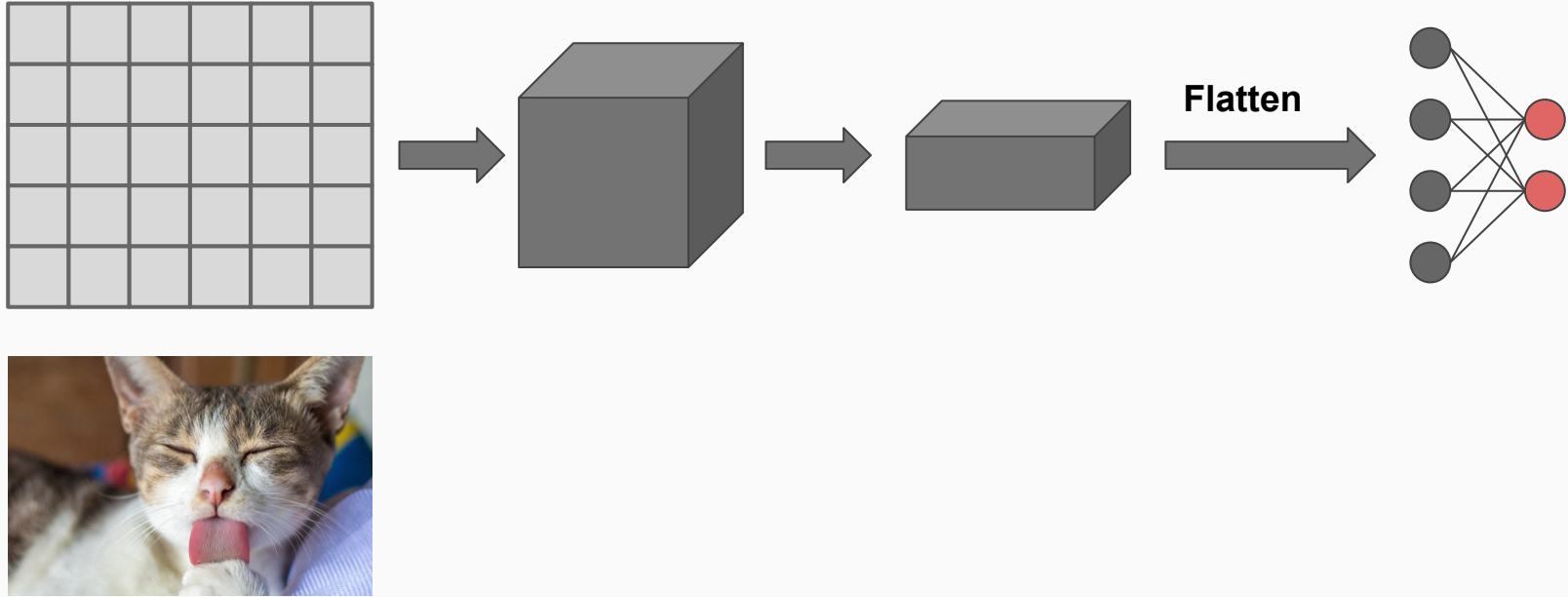
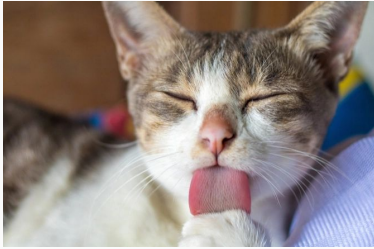
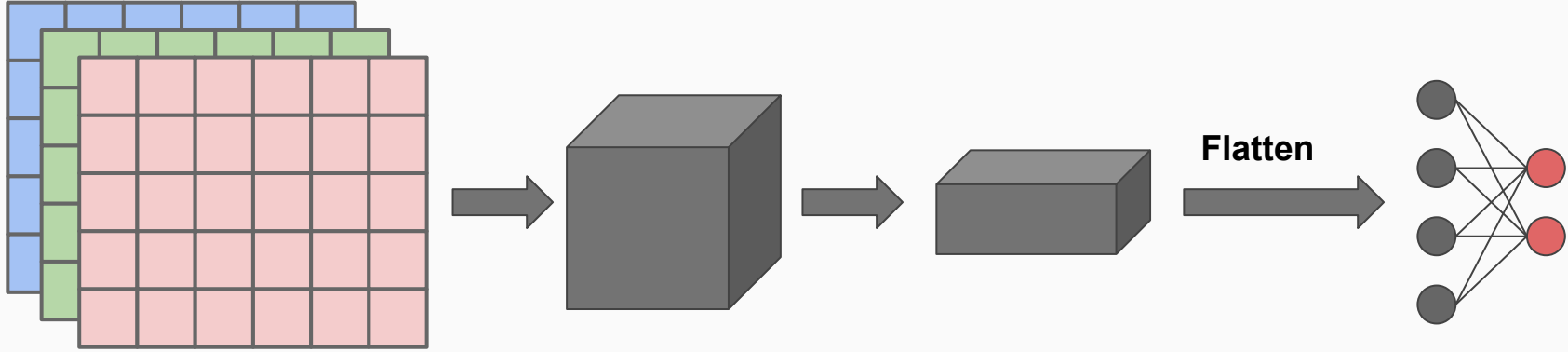
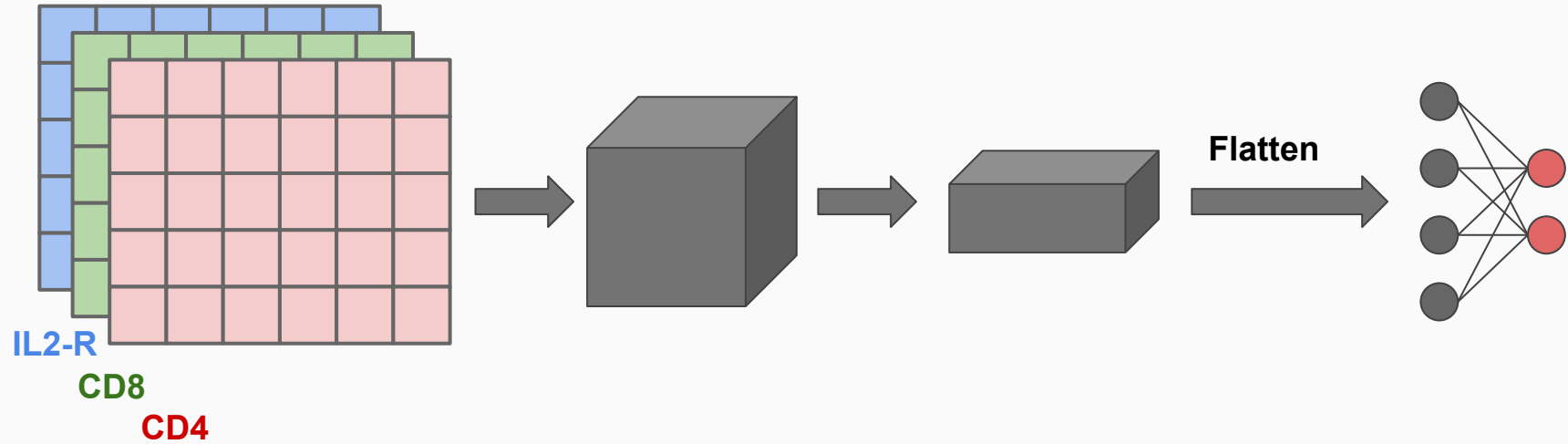


Image -> Convolutional Neural Network (CNN)



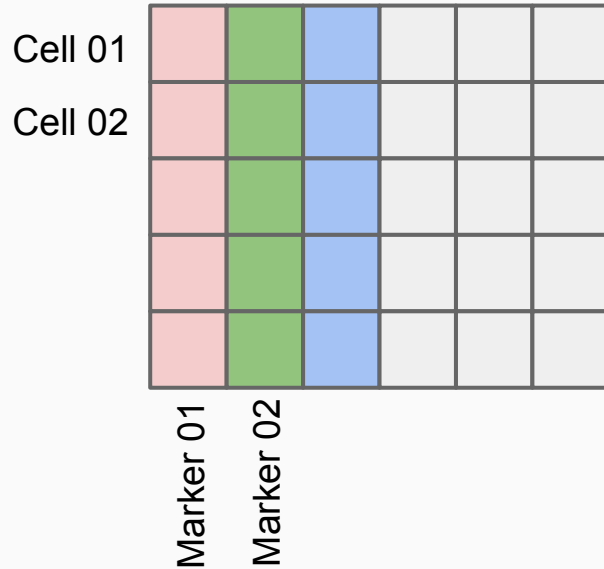
Idea 01: Markers (Genes) as channels



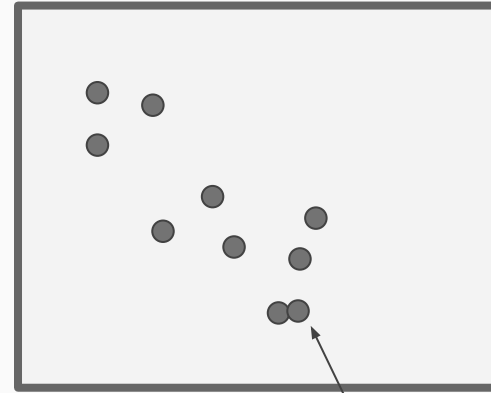
What will be the images?

Idea 02: Generate the plot using t-SNE

Flow Cytometry Data

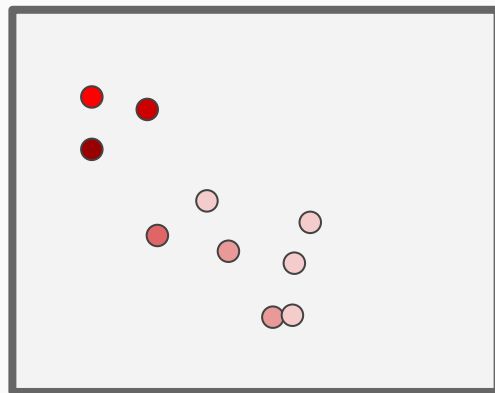
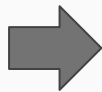
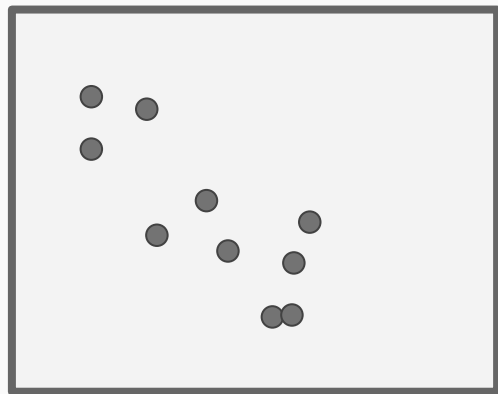


t-SNE

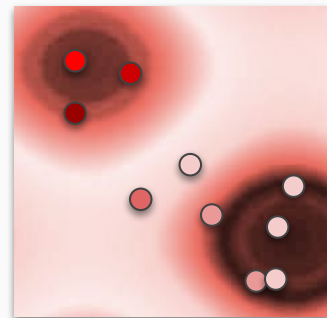
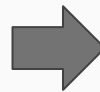


Each point is a cell.

Idea 03: Generate an image from t-SNE plot

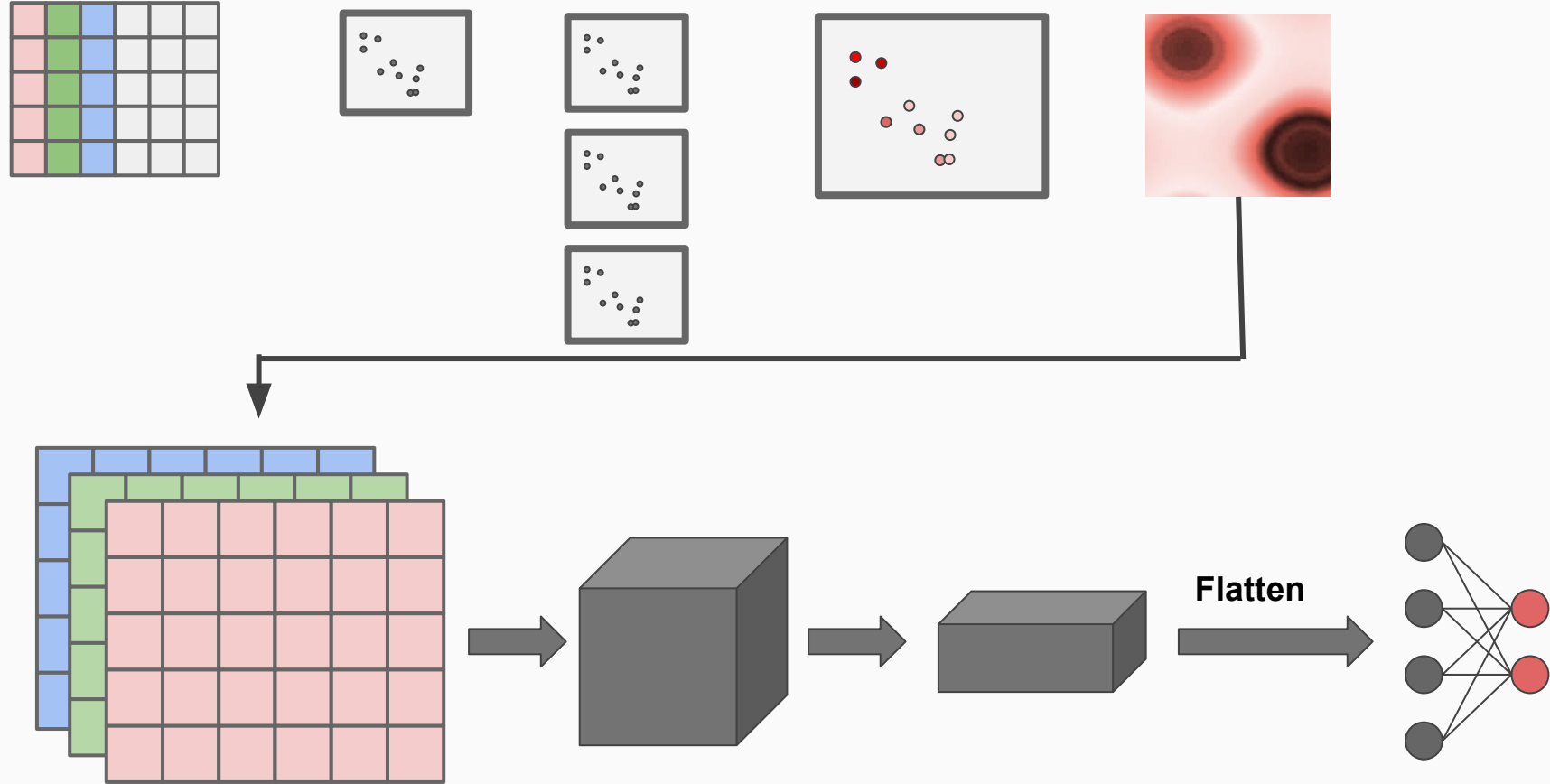


Color by one marker

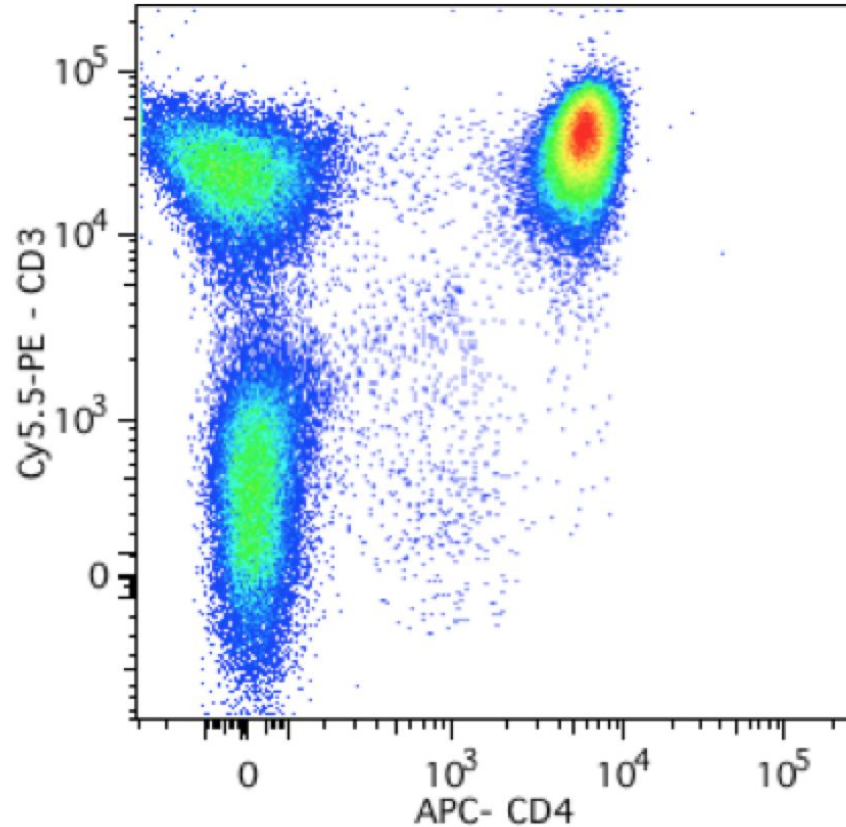


**Color other places by
interpolation**

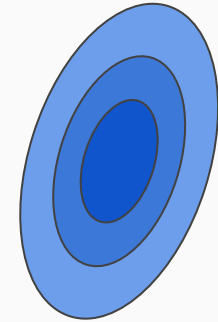
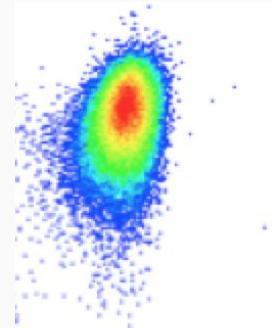
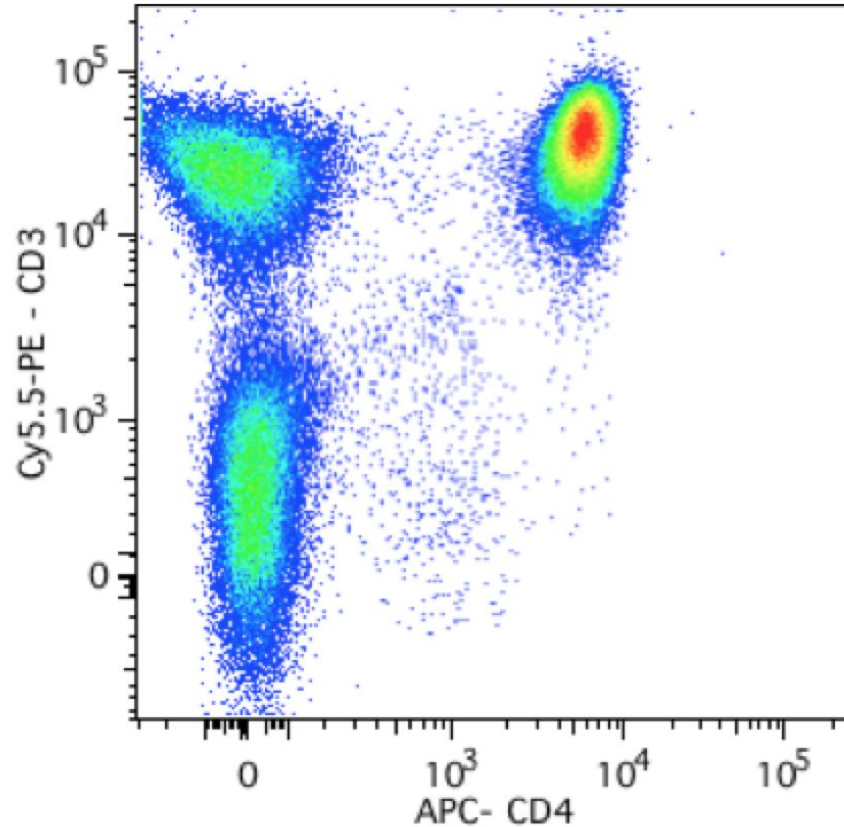
Idea 01 + Idea 02



What does a flow data look like?



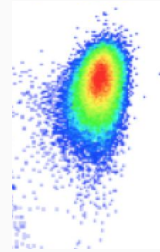
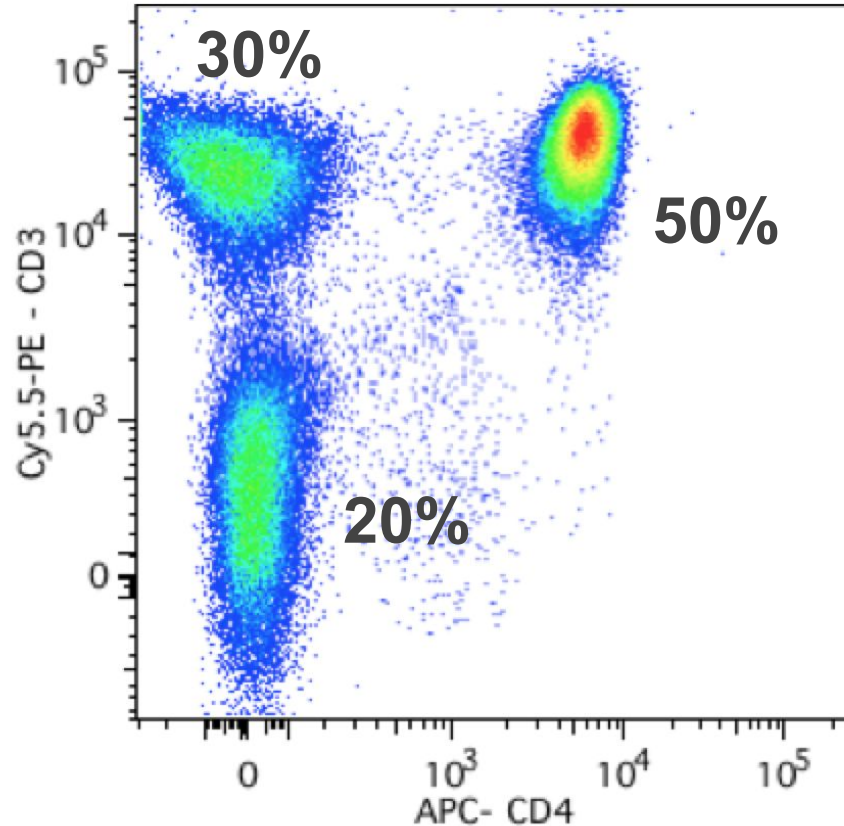
Modeling a flow data



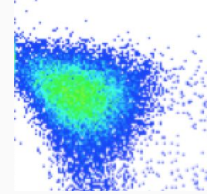
Mean vector (μ)

Variance-Covariance matrix (Σ)

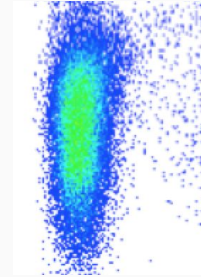
Modeling a flow data



50%



30%



20%

Proportion vector (π)

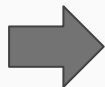
Modeling a flow data

Group 01
(ex: Control)

Mean vector (μ_{01})

Cov Matrix (σ^2_{01})

Proportion vector (π_{01})



Group 02
(ex: Disease)

Mean vector (μ_{02})

Cov Matrix (σ^2_{02})

Proportion vector (π_{02})



**Many samples
from group 01**

**Many samples
from group 02**

Discussion

- Does the tsne-cnn method detect groups with different μ or σ^2 ?
- Does the tsne-cnn method detect groups in real data?
- Could the filter capture meaningful information?

Notes & Discussion

Notes 180320:

- the markers do not have order -> tsne for image
- Goal: giving your blood sample, can we predict your clinical outcome
- It can be applied to different techniques such as the single cell RNA sequencing
- Data augmentation
 - Problem: there are not much samples in a study, therefore, to create more images to train the model, the data augmentation is needed
 - Methods for augmentation
 - Subsetting
 - Different seed of tSNE (similar to rotation of an image)
 - Bootstrapping the existing tSNE plot
- Novel here: using tSNE to create image
- Different CNN methods can be applied to the tSNE image
- Discussion: The feature may not be easy to explain
 - Question: how many features do people usually use?
 - DC gain paper
 - Depend on your dataset
- Discussion: t-SNE is slow
 - Semi-supervised model to gain ????? (what does this mean?)
 - Feed forward gain ?????
 - Transfer learning ?????
 - Try other dim reduction methods (Ex: PCA)
- Discussion: when Subsetting, what if there is only a little difference?
- Discussion: when interpolation, what if there is only a little difference?

Discussion: Autoencoder to perform the dimensional reduction