

Module 3: Clustering and Machine Learning



Instructor: Zach Fox
e-mail: email@lanl.gov



Instructor: Huy Vo
e-mail: Huy.Vo@colostate.edu

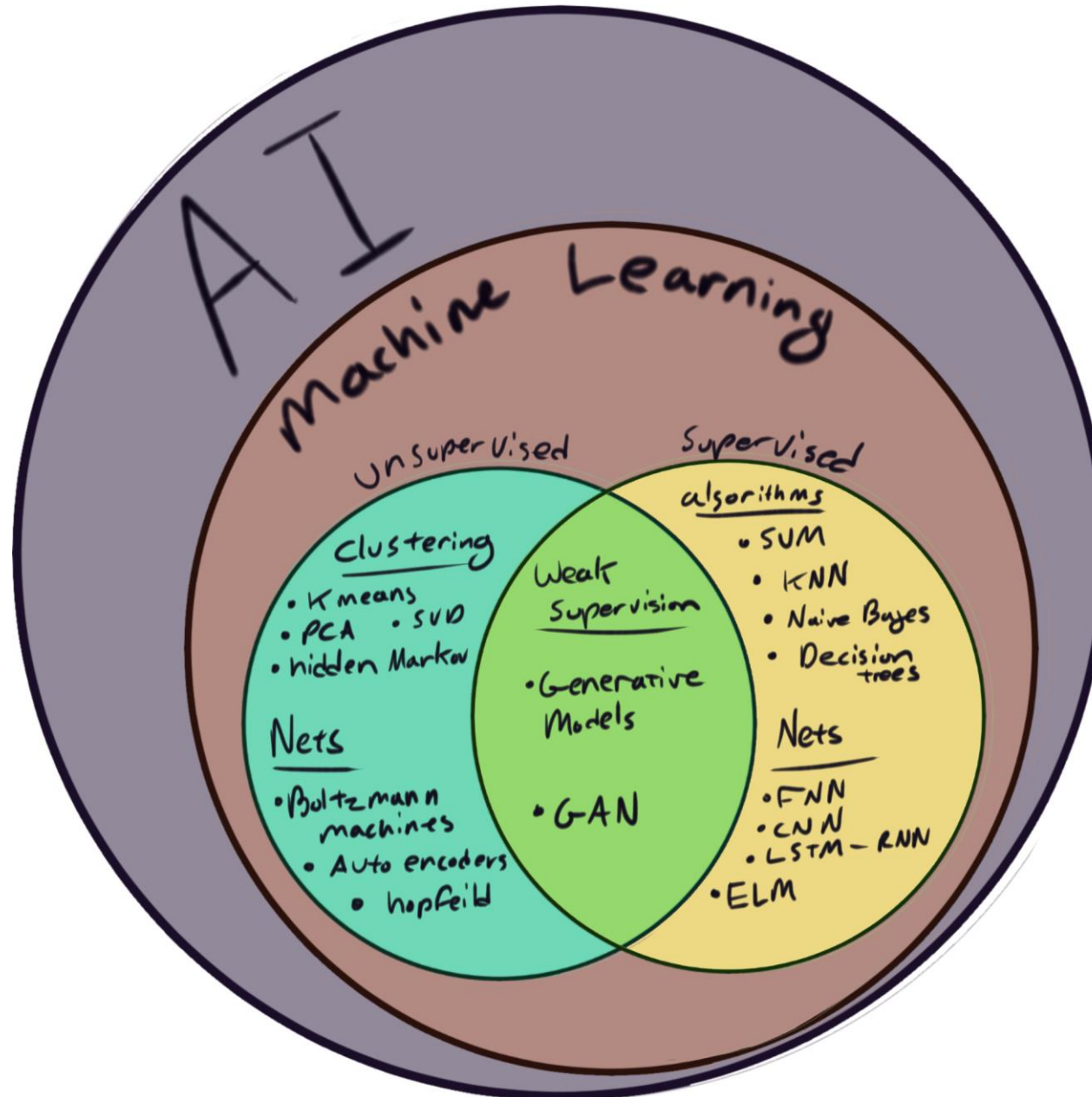


Instructor: William Raymond
e-mail: wsraymon@rams.colostate.edu



Outline

- What is machine learning?
- Unsupervised Learning vs Supervised vs Decision / reinforcement
- Regression vs Classification
- Simple Perceptron
- Neural Networks
- Common Neural Networks
- Activation Functions
- Common Loss functions
- Common Optimizers
- Machine Learning in the context of images
- Ethics



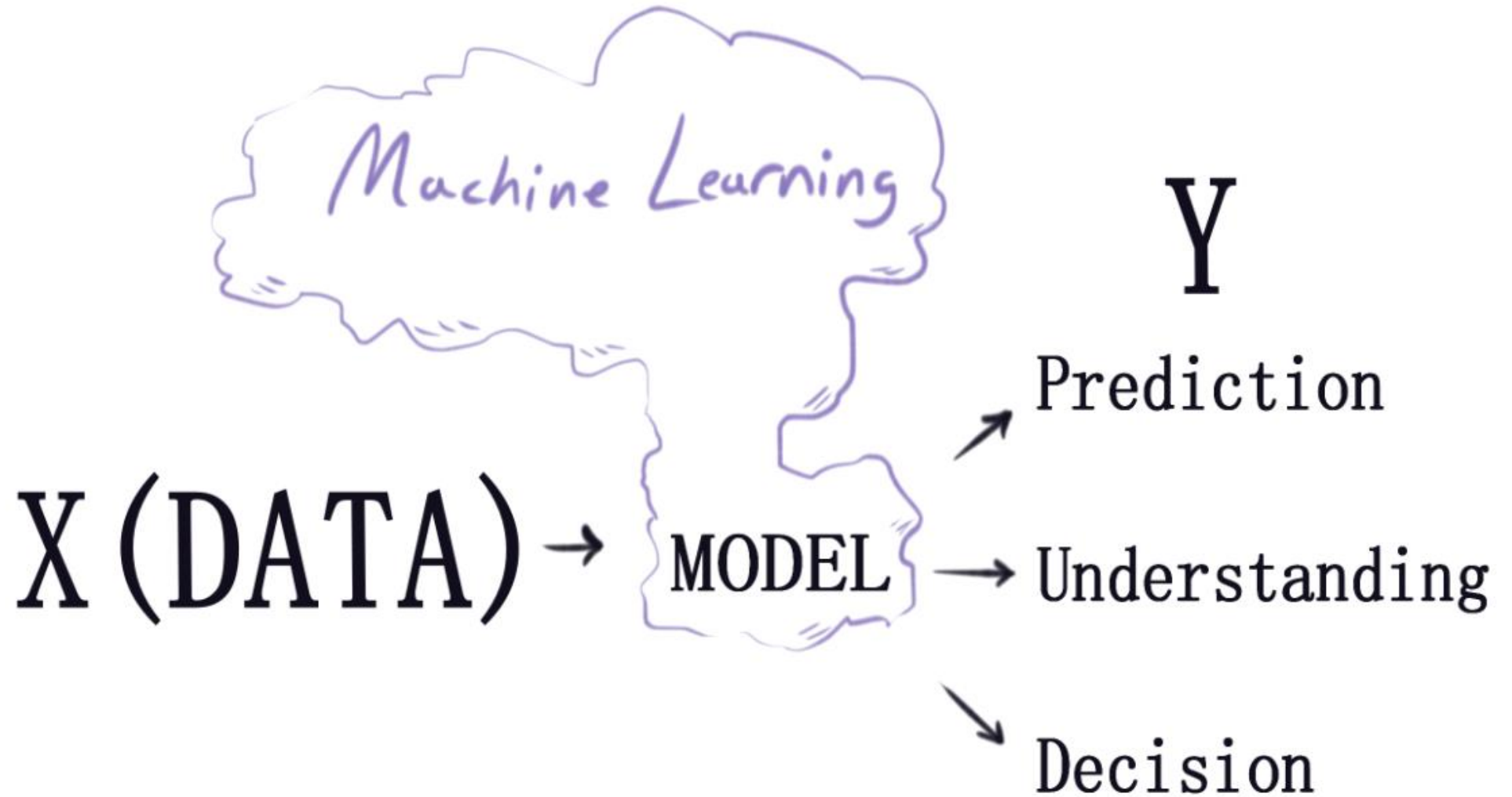
Machine learning is a group of data analysis tools, algorithms, and statistics methods that broadly lies under the umbrella of Artificial Intelligence.

Broadly it can be thought of as any method that aims to produce a model to perform a task (regression or classification) without being explicitly coded.

Machine Learning Models

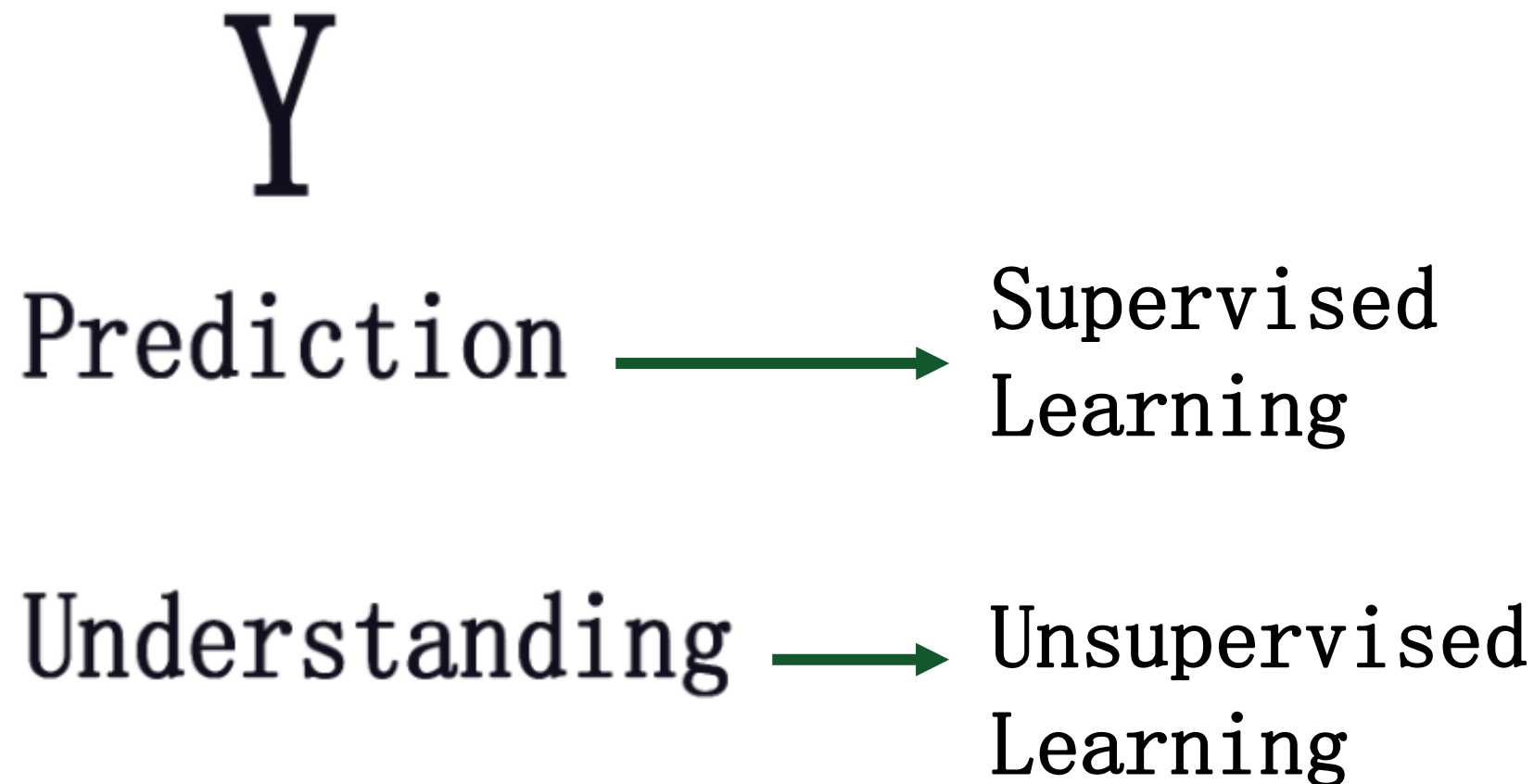
Outline

- What is machine learning?
- Unsupervised Learning vs Supervised vs Decision / reinforcement
- Regression vs Classification
- Simple Perceptron
- Neural Networks
- Common Neural Networks
- Activation Functions
- Common Loss functions
- Common Optimizers
- Machine Learning in the context of images
- Ethics



Outline

- What is machine learning?
- **Unsupervised Learning vs Supervised vs Decision / reinforcement**
- Regression vs Classification
- Simple Perceptron
- Neural Networks
- Common Neural Networks
- Activation Functions
- Common Loss functions
- Common Optimizers
- Machine Learning in the context of images
- Ethics



Outline

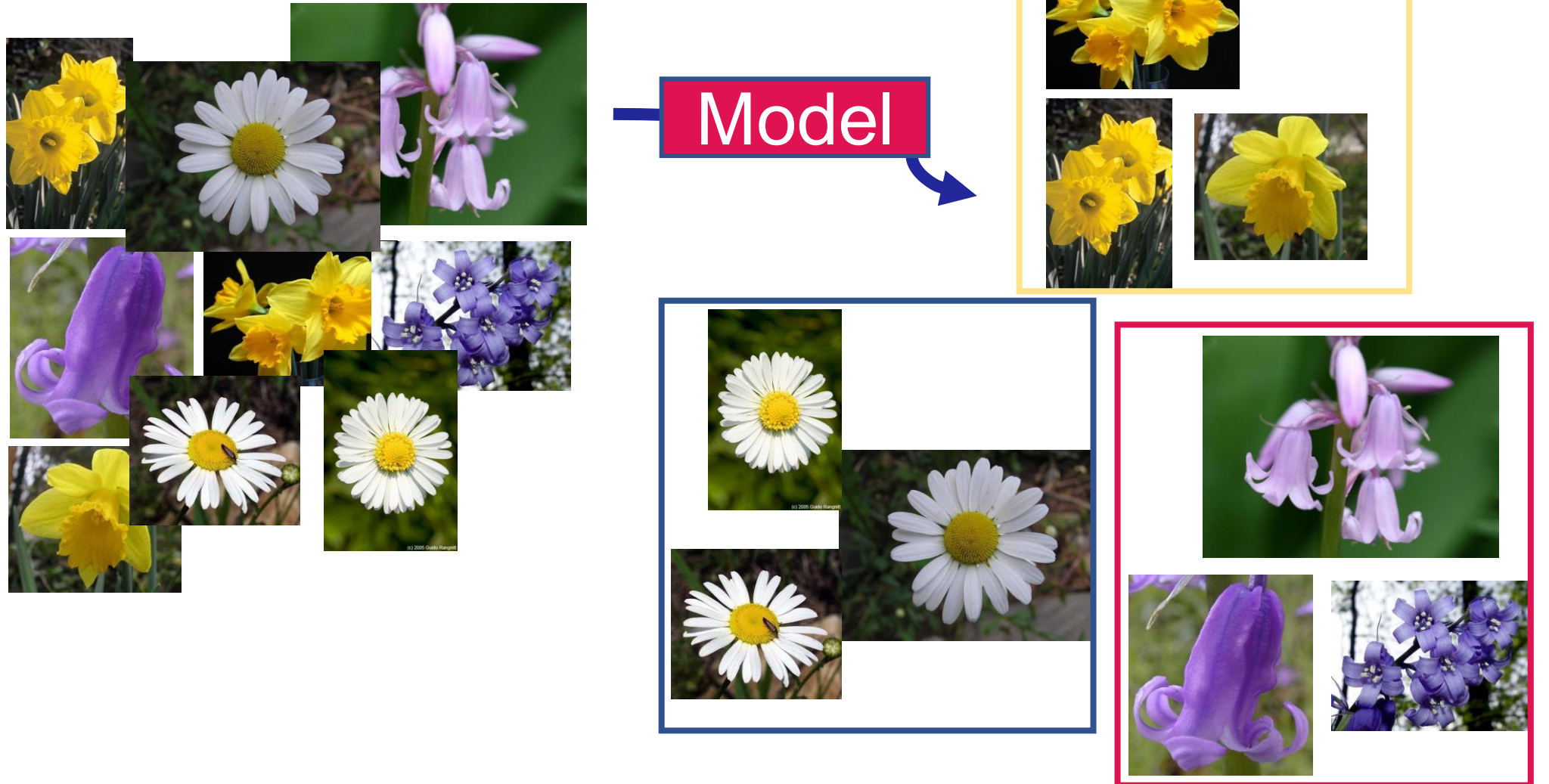
- What is machine learning?
- **Unsupervised Learning vs Supervised vs Decision / reinforcement**
- Regression vs Classification
- Simple Perceptron
- Neural Networks
- Common Neural Networks
- Activation Functions
- Common Loss functions
- Common Optimizers
- Machine Learning in the context of images
- Ethics

Unsupervised Learning

Unsupervised Learning

- Outline**
- What is machine learning?
 - **Unsupervised Learning vs Supervised vs Decision / reinforcement**
 - Regression vs Classification
 - Simple Perceptron
 - Neural Networks
 - Common Neural Networks
 - Activation Functions
 - Common Loss functions
 - Common Optimizers
 - Machine Learning in the context of images
 - Ethics

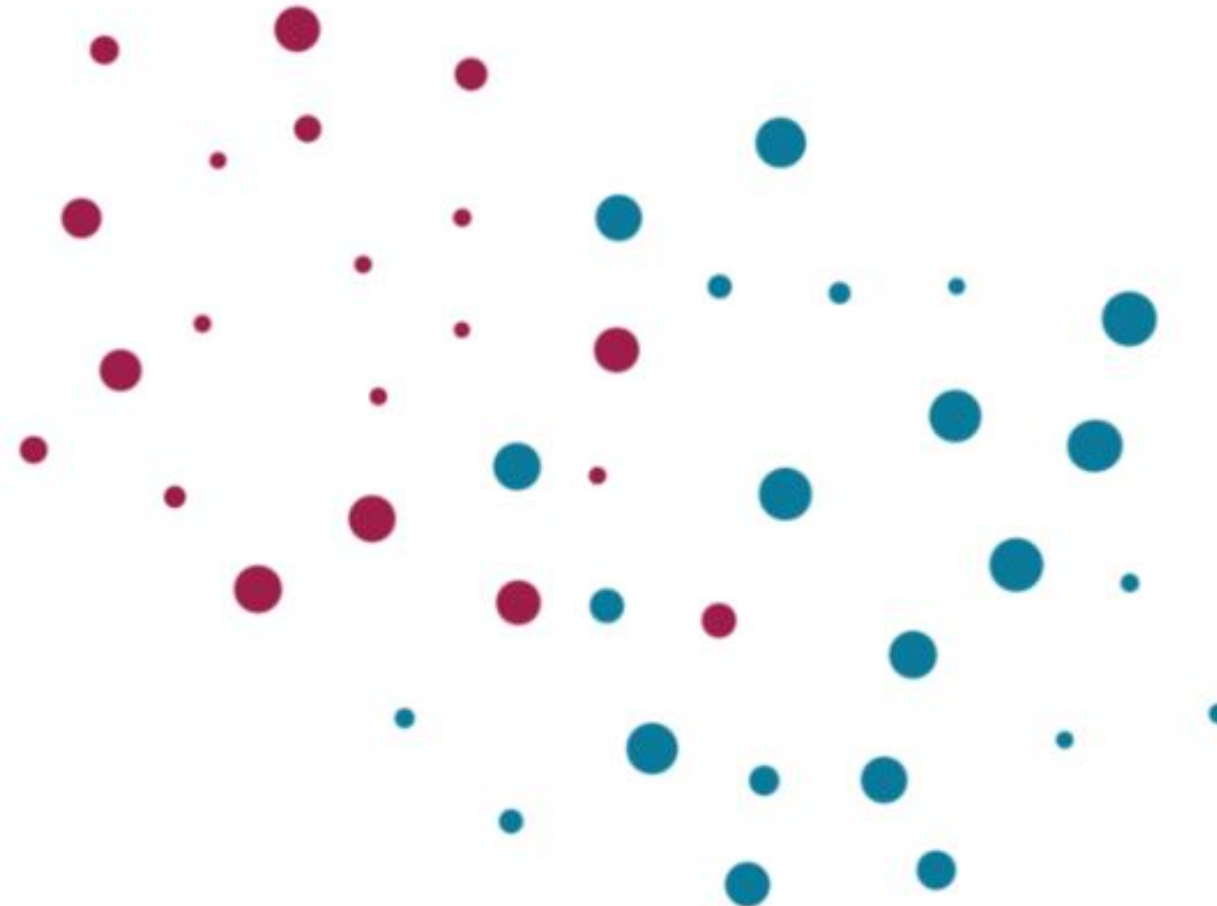
Unsupervised learning is taking a dataset with little to no prior information or labels and giving it order or dimensionality reduction.



Outline

- What is machine learning?
- **Unsupervised Learning vs Supervised vs Decision / reinforcement**
- Regression vs Classification
- Simple Perceptron
- Neural Networks
- Common Neural Networks
- Activation Functions
- Common Loss functions
- Common Optimizers
- Machine Learning in the context of images
- Ethics

$n=2$



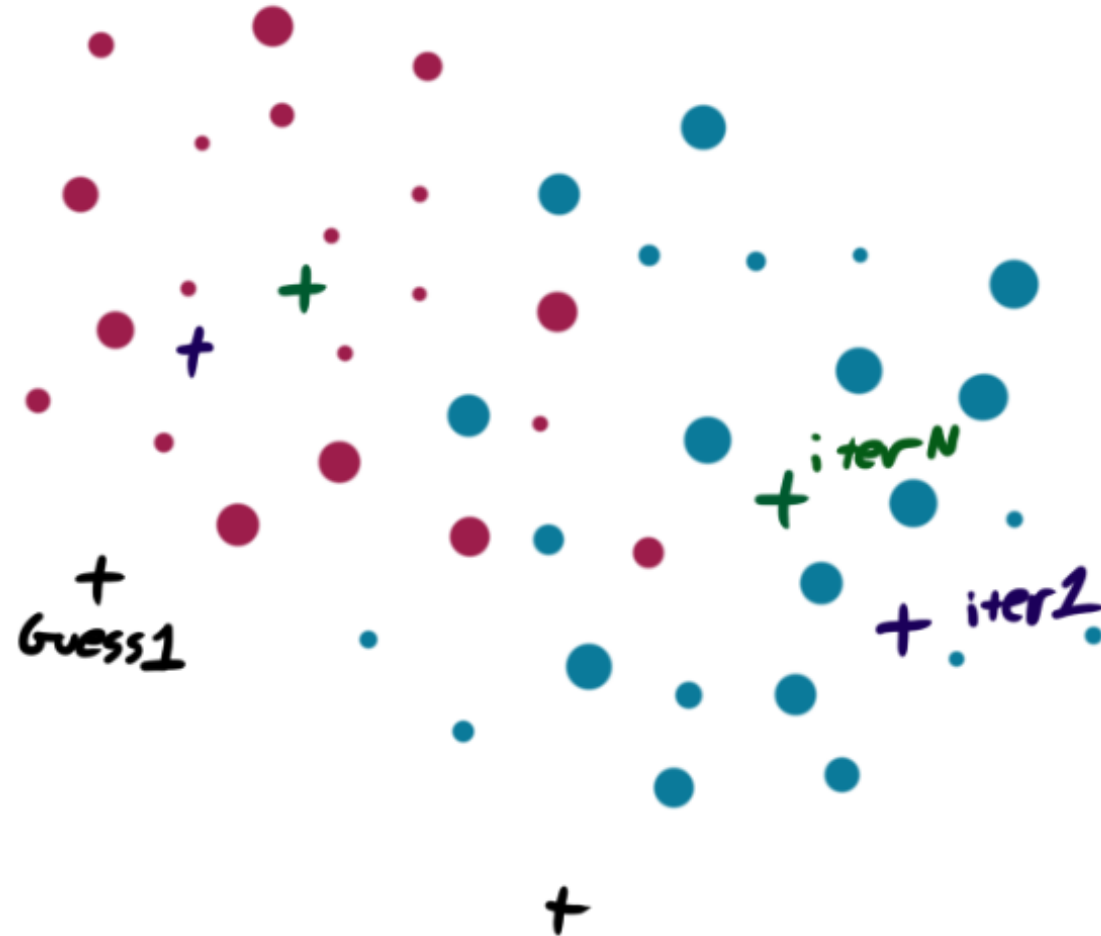
Outline

- What is machine learning?
- **Unsupervised Learning vs Supervised vs Decision / reinforcement**
- Regression vs Classification
- Simple Perceptron
- Neural Networks
- Common Neural Networks
- Activation Functions
- Common Loss functions
- Common Optimizers
- Machine Learning in the context of images
- Ethics

K means

“Giving you n centers, minimize the distance of those centers to its surrounding cluster points”

$n=2$



Clustering

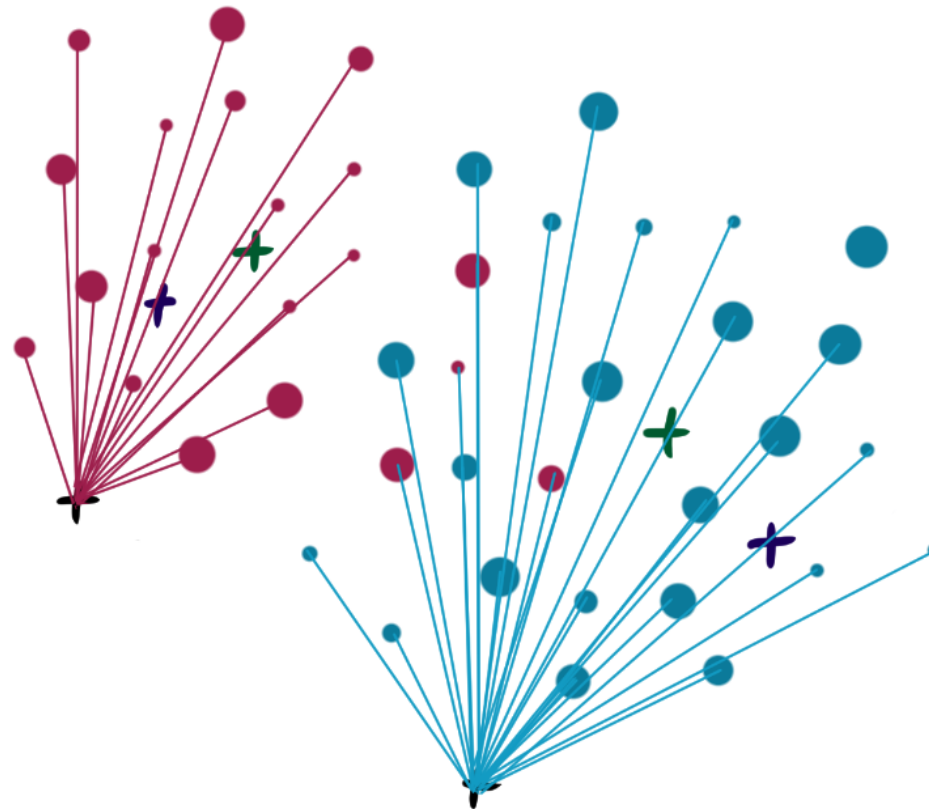
Outline

- What is machine learning?
- **Unsupervised Learning vs Supervised vs Decision / reinforcement**
- Regression vs Classification
- Simple Perceptron
- Neural Networks
- Common Neural Networks
- Activation Functions
- Common Loss functions
- Common Optimizers
- Machine Learning in the context of images
- Ethics

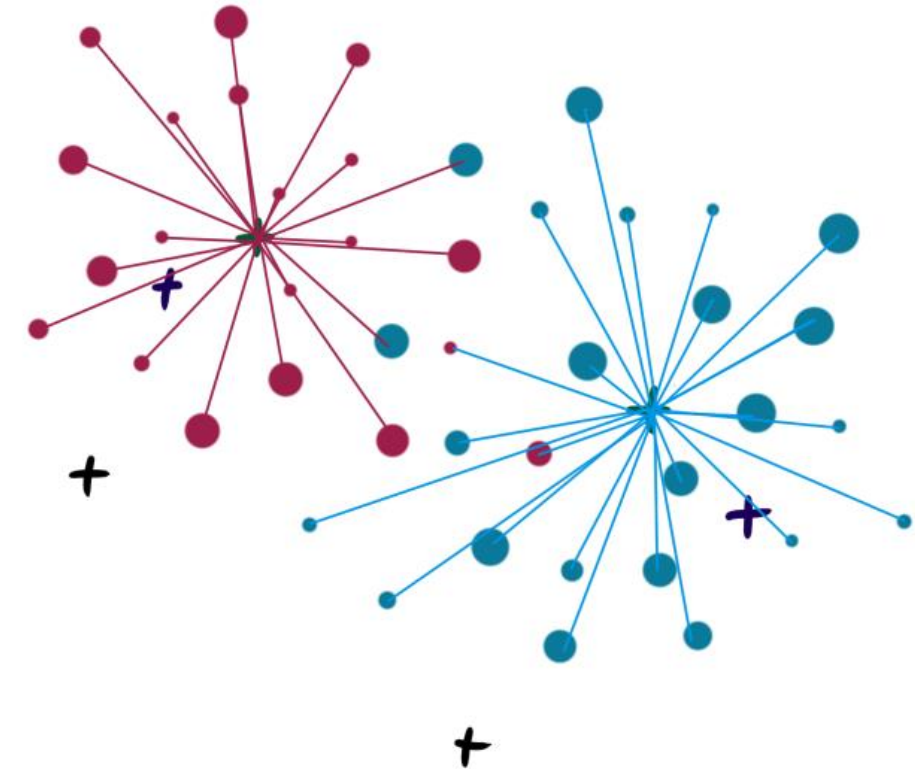
K means

$$\arg \min_{\mathbf{S}} \sum_{i=1}^k \sum_{\mathbf{x} \in S_i} \|\mathbf{x} - \boldsymbol{\mu}_i\|^2$$

Initial Guess



Final Iteration



How are points classified?

Whichever centroid (k-mean) is closest to the point by least squared distance

Clustering

Outline

- What is machine learning?
- **Unsupervised Learning vs Supervised vs Decision / reinforcement**
- Regression vs Classification
- Simple Perceptron
- Neural Networks
- Common Neural Networks
- Activation Functions
- Common Loss functions
- Common Optimizers
- Machine Learning in the context of images
- Ethics

PCA

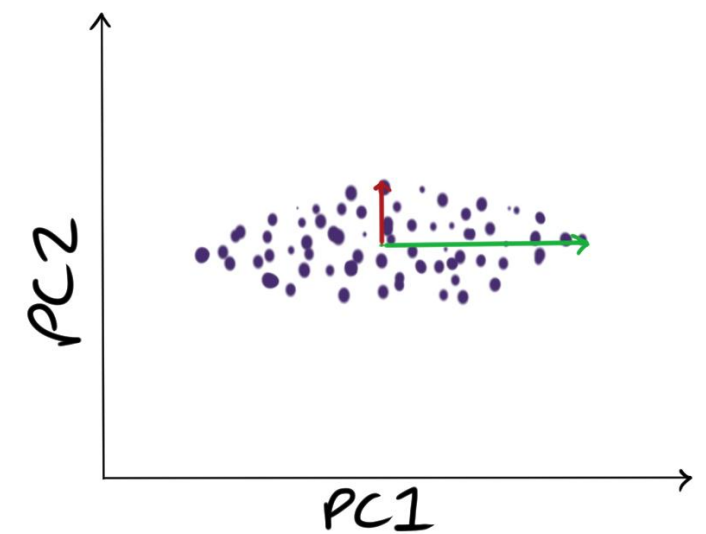
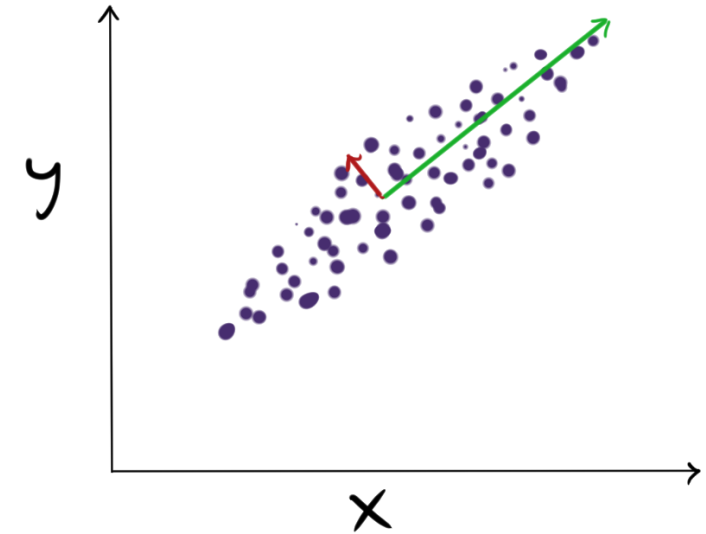
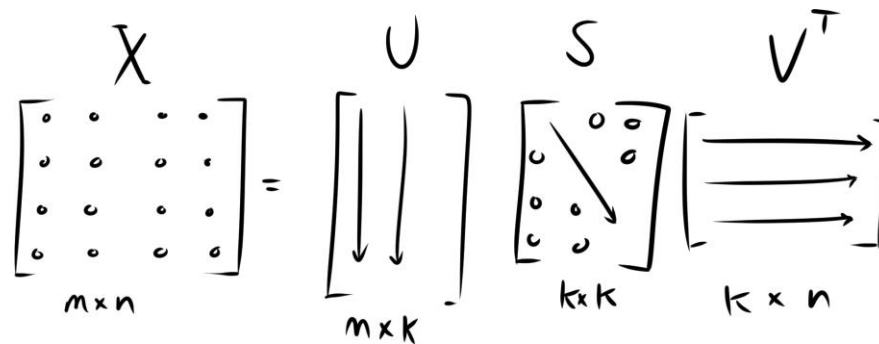
A Transform that changes the coordinates to a system where each coordinate is the least correlated to each other, to a "component."

X, Y, Z, ... \longrightarrow PC1, PC2, PC3

Easiest way to compute is with Singular Value Decomposition (SVD)

$$X = USV'$$

$$\text{PCA} = US \text{ or } XV$$

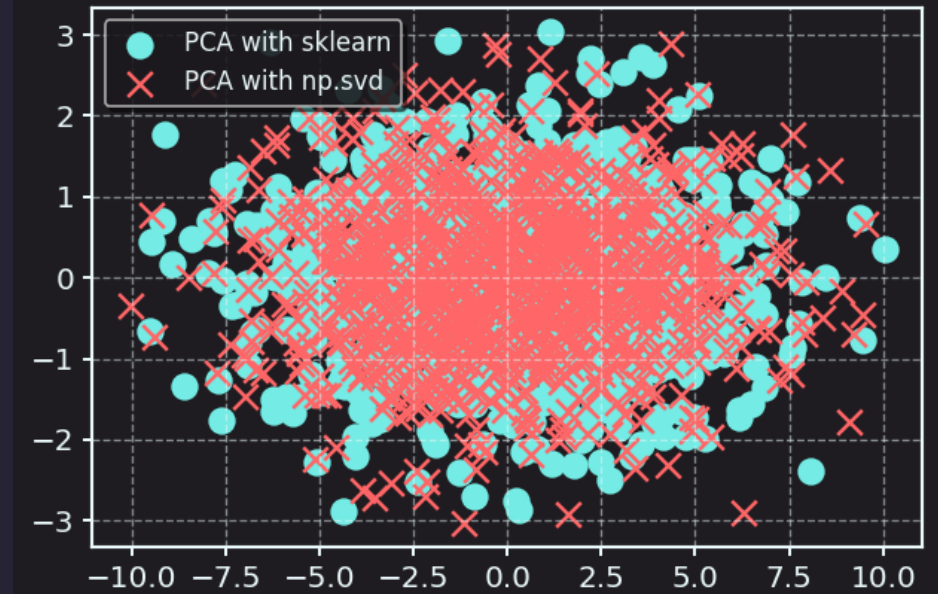
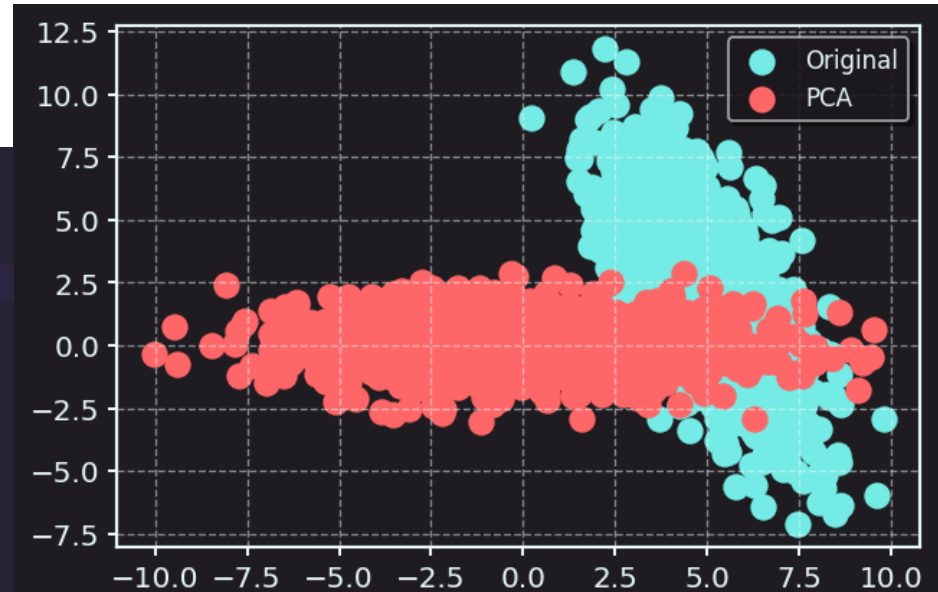


Outline

- What is machine learning?
- **Unsupervised Learning vs Supervised vs Decision / reinforcement**
- Regression vs Classification
- Simple Perceptron
- Neural Networks
- Common Neural Networks
- Activation Functions
- Common Loss functions
- Common Optimizers
- Machine Learning in the context of images
- Ethics

PCA Example in Python

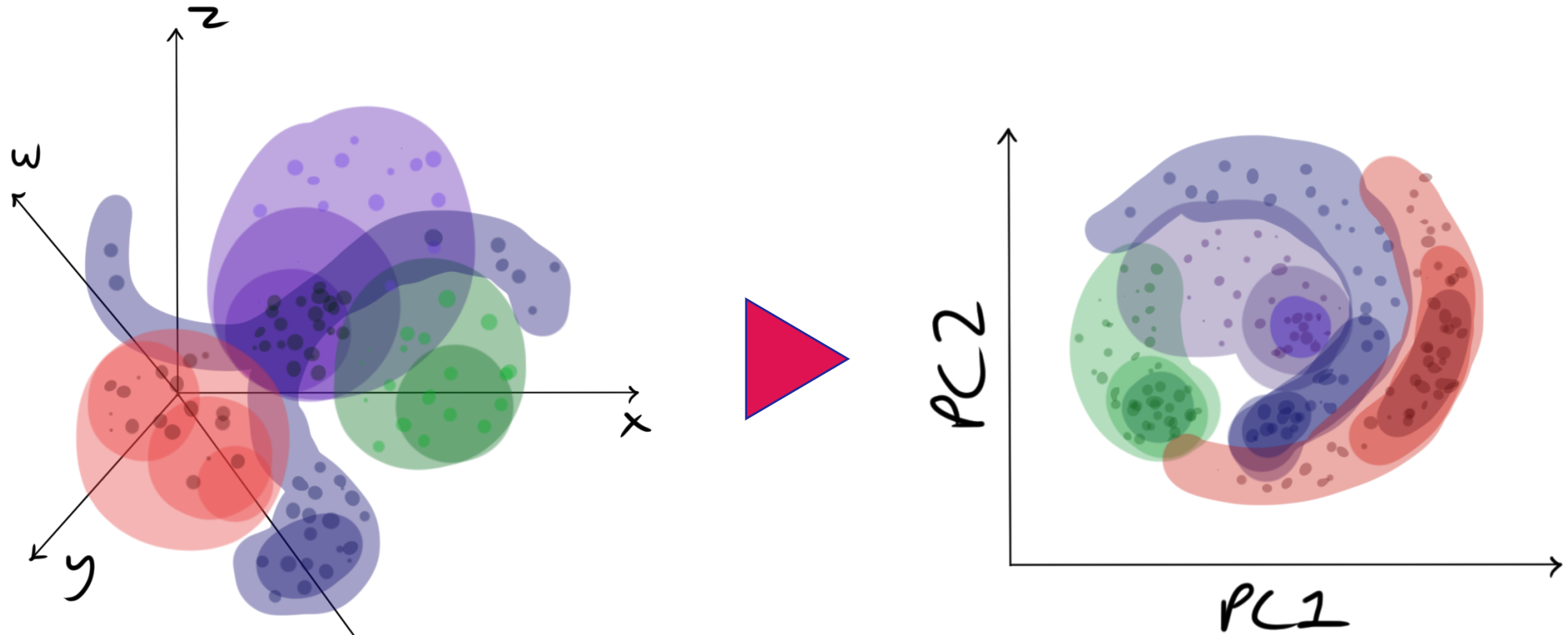
```
1 import numpy as np; import matplotlib.pyplot as plt;
2
3 mean = (5, 2.2)
4 cov = [[2, -3], [-3, 10]]
5 X = np.random.multivariate_normal(mean, cov, (1000))
6
7 plt.scatter(X[:,0],X[:,1])
8
9 U,S,Vt = np.linalg.svd(X- np.mean(X,axis=0)) #decomp of centered X
10
11 U_eig = np.linalg.eig(np.dot(X,X.T))
12 V_eig = np.linalg.eig(np.dot(X.T,X))
13
14
15 print(U.shape)
16 print(S.shape)
17 print(Vt.shape)
18 pca_X = np.dot(X- np.mean(X,axis=0),Vt)
19 plt.scatter(pca_X[:,0],pca_X[:,1])
20 plt.legend(['Original', 'PCA'])
21 plt.figure()
22 from sklearn.decomposition import PCA #comparison with sklearn
23 pca = PCA()
24 pca.fit(X)
25 pca_X_sklearn = pca.transform(X)
26 plt.scatter(pca_X_sklearn[:,0],pca_X_sklearn[:,1])
27 plt.scatter(pca_X[:,0],pca_X[:,1],marker='x')
28 plt.legend(['PCA with sklearn', 'PCA with np.svd'])
```



Outline

- What is machine learning?
- **Unsupervised Learning vs Supervised vs Decision / reinforcement**
- Regression vs Classification
- Simple Perceptron
- Neural Networks
- Common Neural Networks
- Activation Functions
- Common Loss functions
- Common Optimizers
- Machine Learning in the context of images
- Ethics

PCA is commonly used as a simple dimensionality reduction method, where after PCA N dimensional data is plotted along principal components 1 and 2 for 2D visualization or PC1 through 3 for 3D



Outline

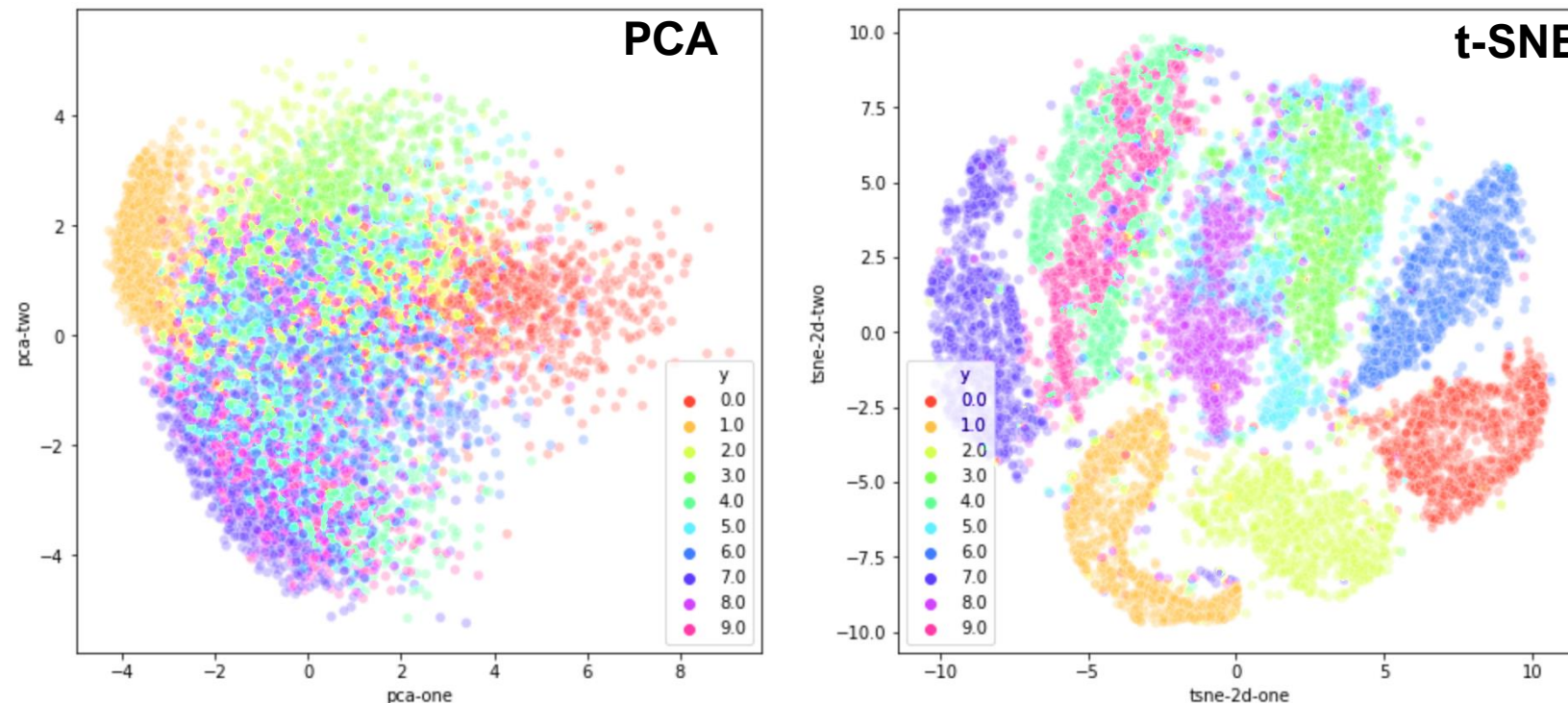
- What is machine learning?
- **Unsupervised Learning vs Supervised vs Decision / reinforcement**
- Regression vs Classification
- Simple Perceptron
- Neural Networks
- Common Neural Networks
- Activation Functions
- Common Loss functions
- Common Optimizers
- Machine Learning in the context of images
- Ethics

t-SNE

Utilized widely in bioinformatics and genomics

1. Calculate probability that points are neighbors,
2. Calculate or learn a dimensional map that keeps this neighbor probability

MNIST Digits

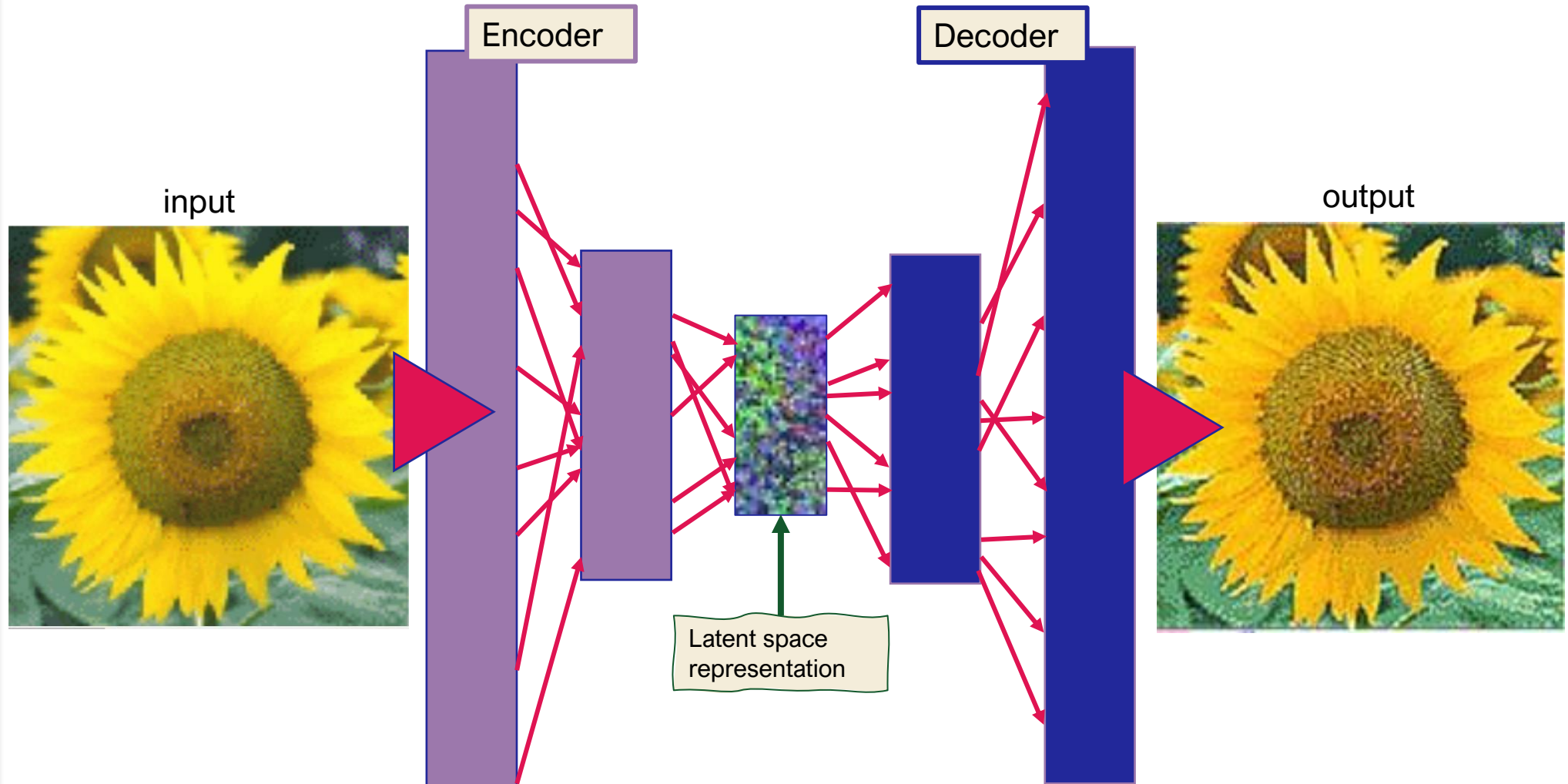


"Clustering"

Outline

- What is machine learning?
- **Unsupervised Learning vs Supervised vs Decision / reinforcement**
- Regression vs Classification
- Simple Perceptron
- Neural Networks
- Common Neural Networks
- Activation Functions
- Common Loss functions
- Common Optimizers
- Machine Learning in the context of images
- Ethics

Autoencoders are a class of neural network that learn a “compression” or latent space representation of information – they are trained on a loss function of how different the input and output information is – typically MSE.



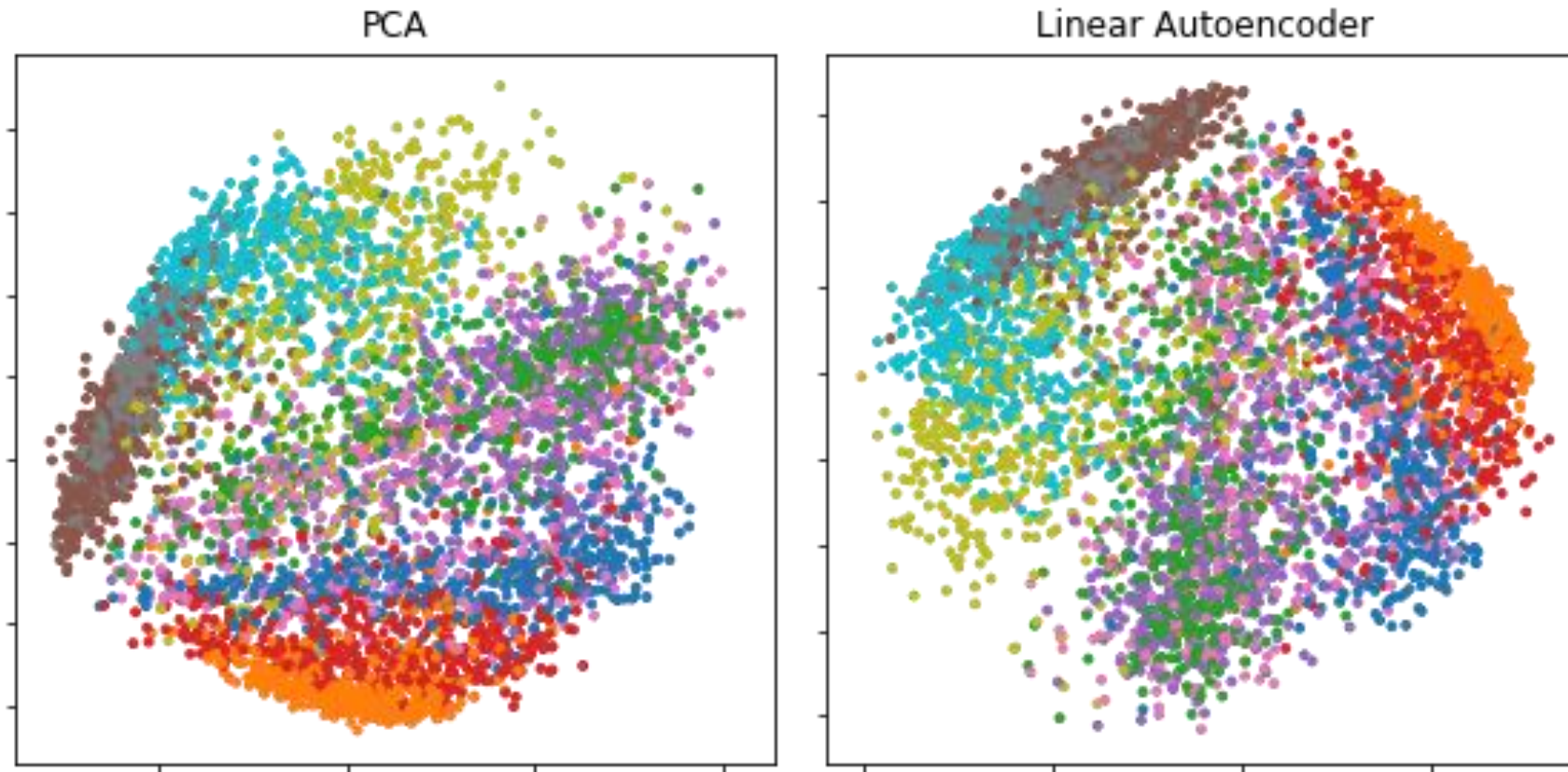
Outline

- What is machine learning?
- **Unsupervised Learning vs Supervised vs Decision / reinforcement**
- Regression vs Classification
- Simple Perceptron
- Neural Networks
- Common Neural Networks
- Activation Functions
- Common Loss functions
- Common Optimizers
- Machine Learning in the context of images
- Ethics

Fashion MNIST principal components vs Linear Autoencoder latent space projection

The Latent space representation learned by the autoencoder can be used to cluster or organize datasets

it's a lower dimension representation of information contained in the dataset



By Michela Massi - Own work, CC BY-SA 4.0, <https://commons.wikimedia.org/w/index.php?curid=80152034>

Latent Space Aside

Outline

- What is machine learning?
- **Unsupervised Learning vs Supervised vs Decision / reinforcement**
- Regression vs Classification
- Simple Perceptron
- Neural Networks
- Common Neural Networks
- Activation Functions
- Common Loss functions
- Common Optimizers
- Machine Learning in the context of images
- Ethics

Input
128x128x3



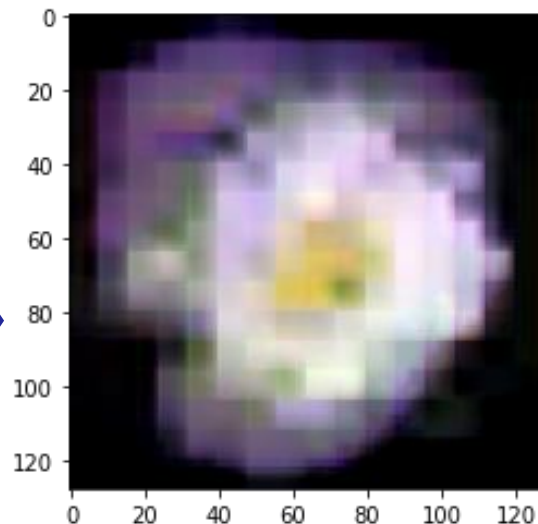
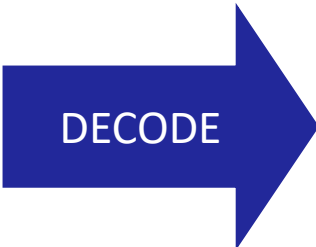
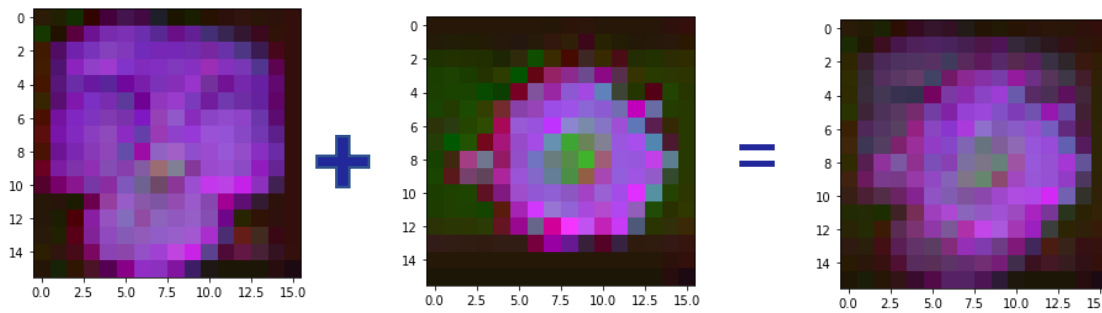
Latent
16x16x3



Output
128x128x3



You can mix latent representations and decode for style mixtures!



Clustering in Python

Outline

- What is machine learning?
- **Unsupervised Learning vs Supervised vs Decision / reinforcement**
- Regression vs Classification
- Simple Perceptron
- Neural Networks
- Common Neural Networks
- Activation Functions
- Common Loss functions
- Common Optimizers
- Machine Learning in the context of images
- Ethics



Method name	Parameters	Scalability	Usecase	Geometry (metric used)
K-Means	number of clusters	Very large <code>n_samples</code> , medium <code>n_clusters</code> with <code>MiniBatch</code> code	General-purpose, even cluster size, flat geometry, not too many clusters, inductive	Distances between points
Affinity propagation	damping, sample preference	Not scalable with <code>n_samples</code>	Many clusters, uneven cluster size, non-flat geometry, inductive	Graph distance (e.g. nearest-neighbor graph)
Mean-shift	bandwidth	Not scalable with <code>n_samples</code>	Many clusters, uneven cluster size, non-flat geometry, inductive	Distances between points
Spectral clustering	number of clusters	Medium <code>n_samples</code> , small <code>n_clusters</code>	Few clusters, even cluster size, non-flat geometry, transductive	Graph distance (e.g. nearest-neighbor graph)
Ward hierarchical clustering	number of clusters or distance threshold	Large <code>n_samples</code> and <code>n_clusters</code>	Many clusters, possibly connectivity constraints, transductive	Distances between points
Agglomerative clustering	number of clusters or distance threshold, linkage type, distance	Large <code>n_samples</code> and <code>n_clusters</code>	Many clusters, possibly connectivity constraints, non-Euclidean distances, transductive	Any pairwise distance
DBSCAN	neighborhood size	Very large <code>n_samples</code> , medium <code>n_clusters</code>	Non-flat geometry, uneven cluster sizes, transductive	Distances between nearest points
OPTICS	minimum cluster membership	Very large <code>n_samples</code> , large <code>n_clusters</code>	Non-flat geometry, uneven cluster sizes, variable cluster density, transductive	Distances between points
Gaussian mixtures	many	Not scalable	Flat geometry, good for density estimation, inductive	Mahalanobis distances to centers
BIRCH	branching factor, threshold, optional global clusterer.	Large <code>n_clusters</code> and <code>n_samples</code>	Large dataset, outlier removal, data reduction, inductive	Euclidean distance between points

<https://scikit-learn.org/stable/modules/clustering.html>

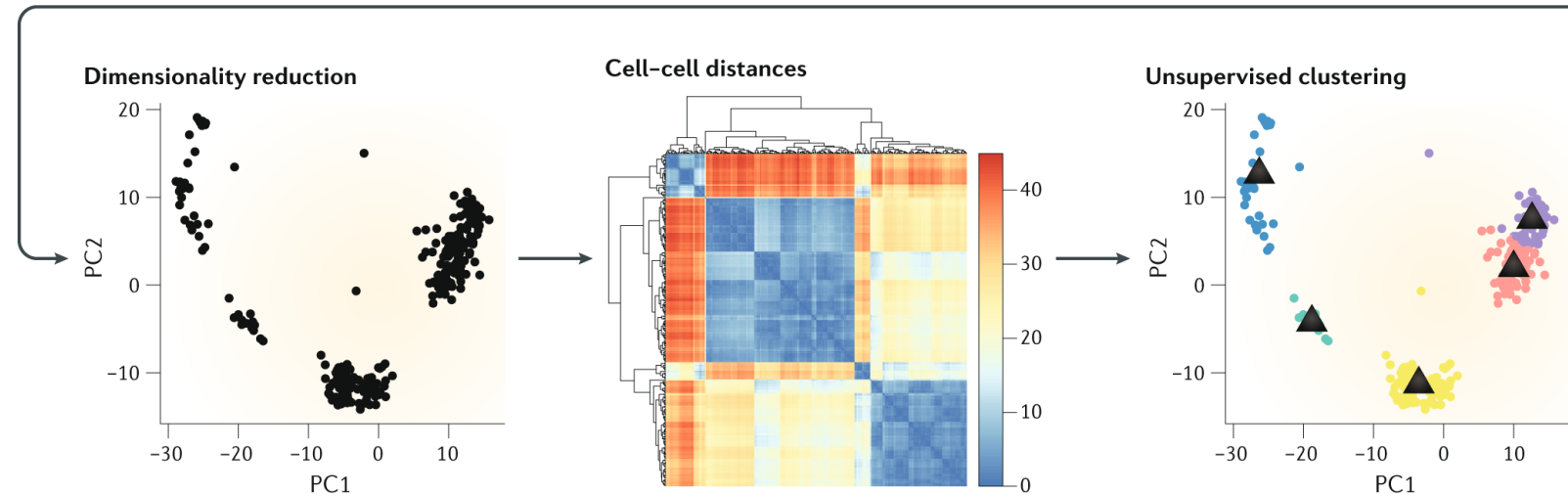
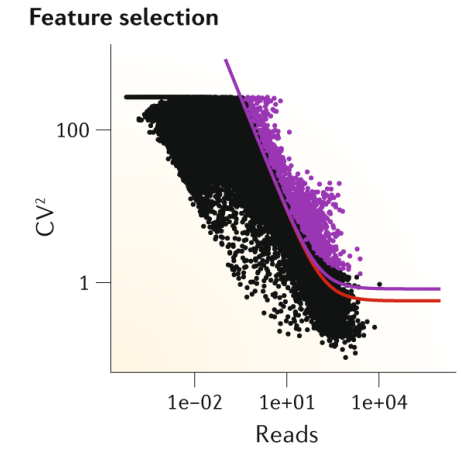
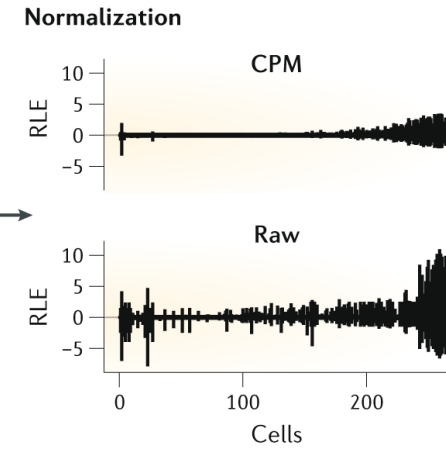
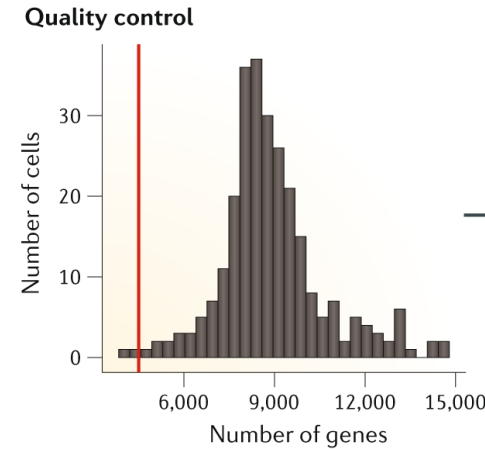
Unsupervised learning in Biological Contexts

Outline

- What is machine learning?
- **Unsupervised Learning vs Supervised vs Decision / reinforcement**
- Regression vs Classification
- Simple Perceptron
- Neural Networks
- Common Neural Networks
- Activation Functions
- Common Loss functions
- Common Optimizers
- Machine Learning in the context of images
- Ethics

Bioinformatics

- Codon bias
- Proteomics / Transcriptomics
- Epigenomics



Challenges in unsupervised clustering of single-cell RNA-seq data. Kiselev et al. Nature Reviews Genetics 2019. <https://www.nature.com/articles/s41576-018-0088-9>

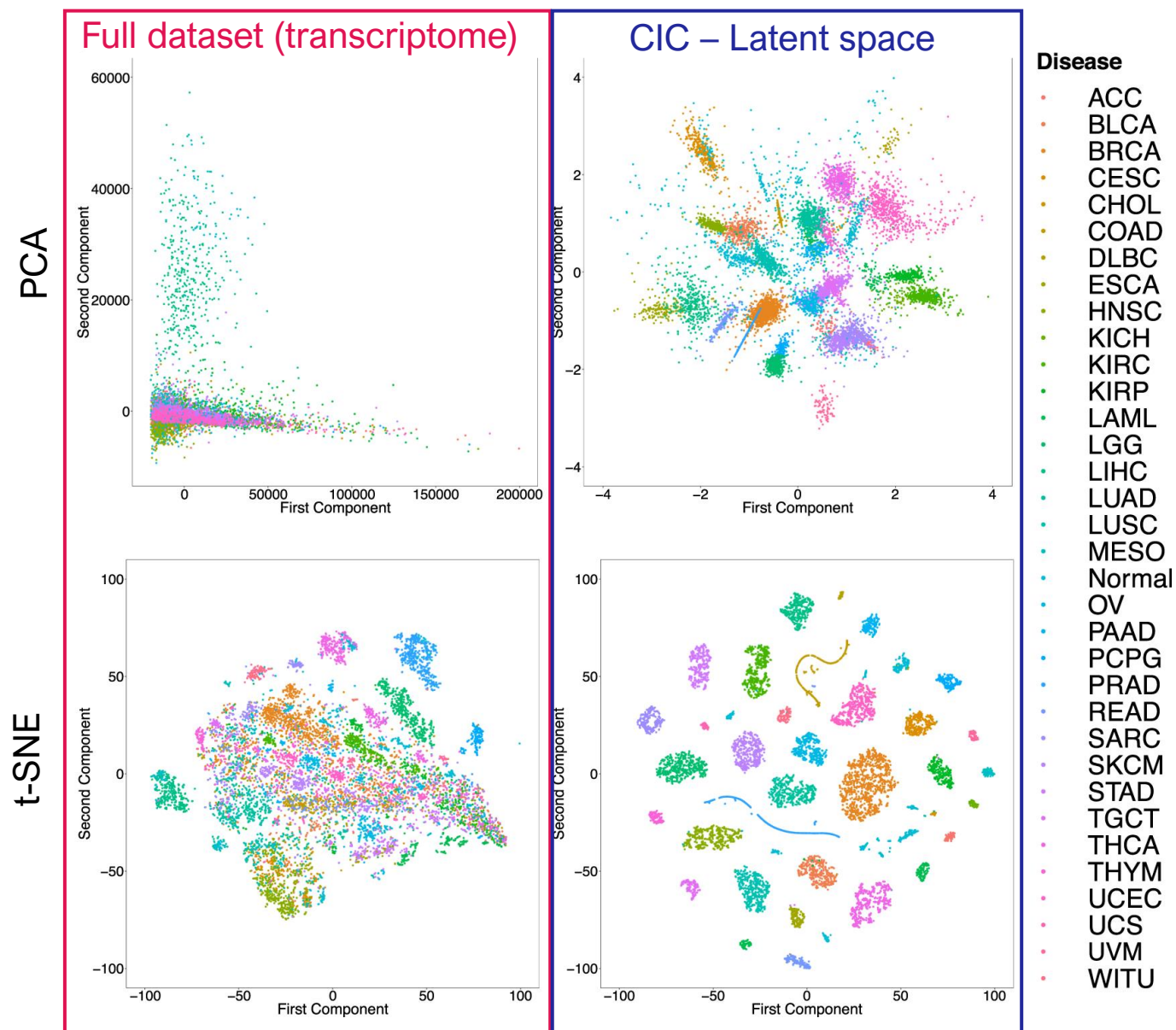
Outline

- What is machine learning?
- **Unsupervised Learning vs Supervised vs Decision / reinforcement**
- Regression vs Classification
- Simple Perceptron
- Neural Networks
- Common Neural Networks
- Activation Functions
- Common Loss functions
- Common Optimizers
- Machine Learning in the context of images
- Ethics

Autoencoders are often used to produce latent space representations for use in later deep learning classification models.

Using latent space representations can help denoise large datasets (such as RNA-seq provides) and provide a smaller dimensional dataset – speeding up training and hyperparameter searches for models.

In the figure to the right a learned “Cell Identity Code” (8D latent space representation) allows for clearer clustering of mRNA profiles of cancer biopsies than the original dataset.



Outline

- What is machine learning?
- **Unsupervised Learning vs Supervised vs Decision / reinforcement**
- Regression vs Classification
- Simple Perceptron
- Neural Networks
- Common Neural Networks
- Activation Functions
- Common Loss functions
- Common Optimizers
- Machine Learning in the context of images
- Ethics

Supervised Learning

Supervised Learning

Outline

- What is machine learning?
- **Unsupervised Learning vs Supervised vs Decision / reinforcement**
- Regression vs Classification
- Simple Perceptron
- Neural Networks
- Common Neural Networks
- Activation Functions
- Common Loss functions
- Common Optimizers
- Machine Learning in the context of images
- Ethics



Cat (label 0)



Dog (label 1)

Supervised learning is when each data point within a dataset has an associated label or target for the machine learning model to match.

Model



Outline

- What is machine learning?
- **Unsupervised Learning vs Supervised vs Decision / reinforcement**
- Regression vs Classification
- Simple Perceptron
- Neural Networks
- Common Neural Networks
- Activation Functions
- Common Loss functions
- Common Optimizers
- Machine Learning in the context of images
- Ethics

Examples:

Melanoma Classification

- > 5000 papers
- Predominantly CNN based

Protein Folding

- Holy grail of machine learning
- Sequence to structure
- CASP

Cardiac Pathology Diagnosis

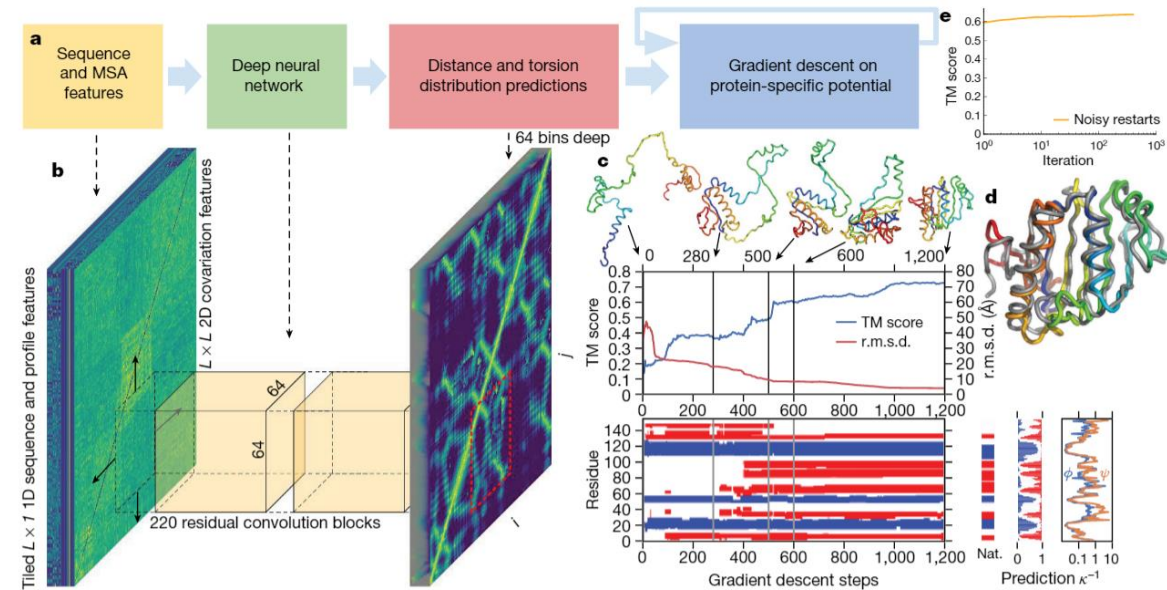
- EKG Signal to detect defects

Disease state (Cancer) Diagnosis

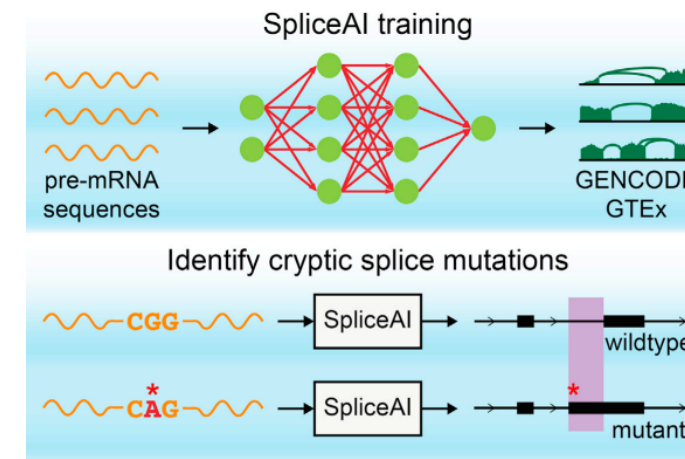
- Label Global mRNA expression profiles to train classification of tumors
- Proteomics + RNA-seq

Genomics

- Splice site recognition



Alpha Fold - Improved protein structure prediction using potentials from deep learning. Senior et al. eLife Nature 2020. DOI: <https://doi.org/10.1038/s41586-019-1923-7>

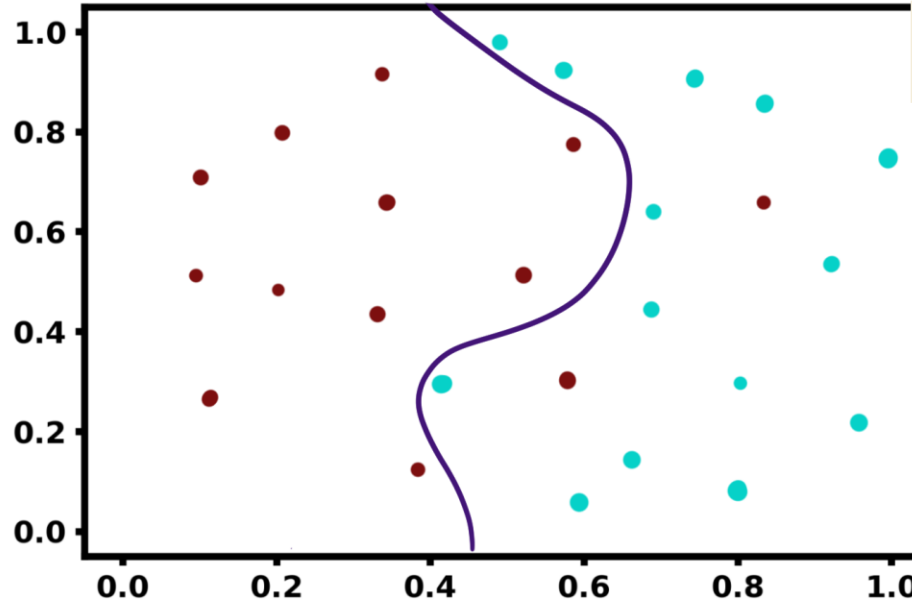


Predicting Splicing from Primary Sequence with Deep Learning. Jaganathan et al. Cell 2019. DOI: <https://doi.org/10.1016/j.cell.2018.12.015>

Regression vs Classification

Outline

- What is machine learning?
- Unsupervised Learning vs Supervised vs Decision / reinforcement
- **Regression vs Classification**
- Simple Perceptron
- Neural Networks
- Common Neural Networks
- Activation Functions
- Common Loss functions
- Common Optimizers
- Machine Learning in the context of images
- Ethics

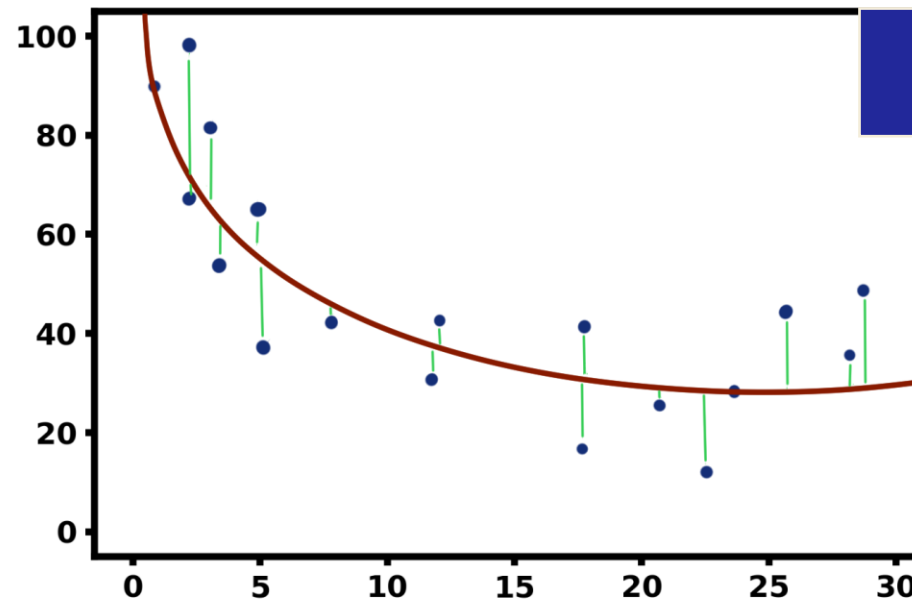


Discrete labels (y)

“Given these features or photos, what is the house color?”

Y: [p(white), p(black), p(red), p(blue), Etc]

Other examples:
house or apartment?
City or rural?



Continuous output (y)

“Given these features, what is the market price of this house?”

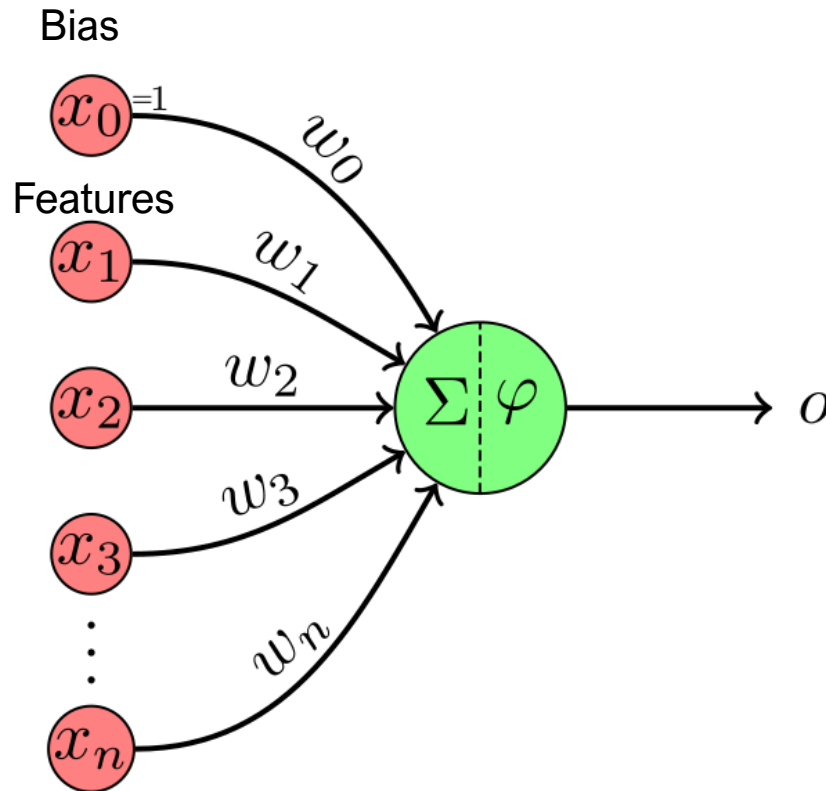
Y: 0 – X \$

Other examples:
What’s the square footage?
How big is the yard?
How long will it take to sell?

Simple Perceptron

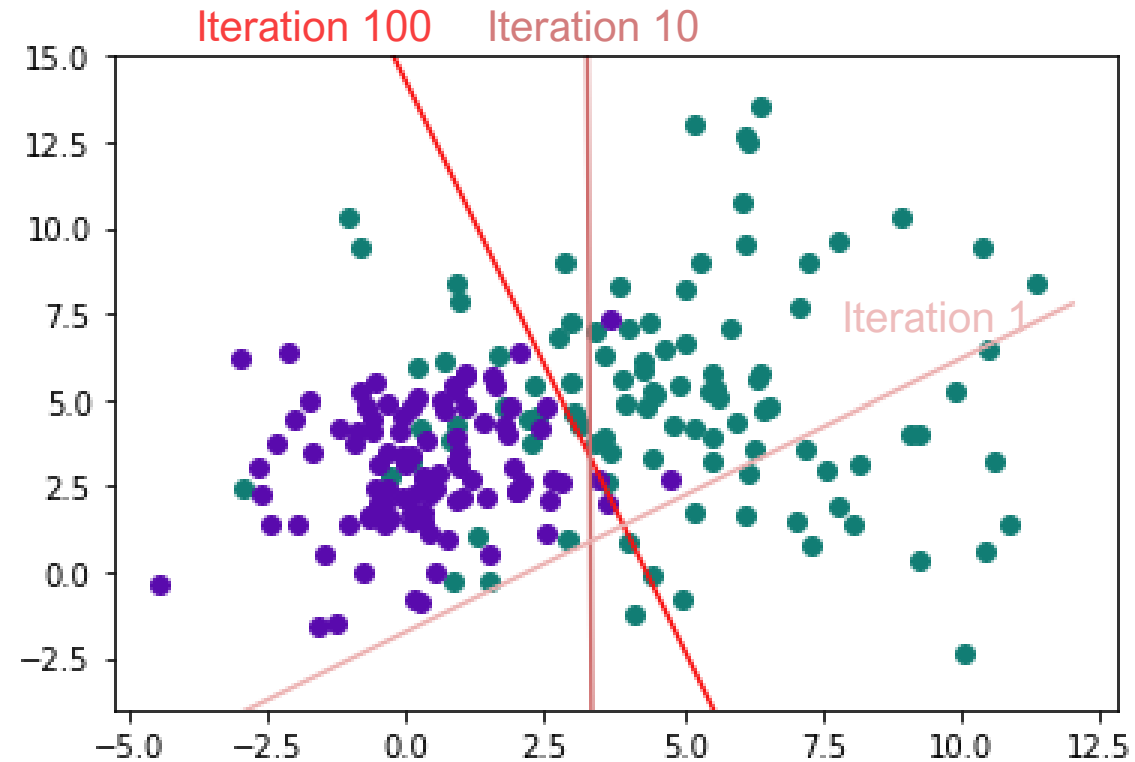
Outline

- What is machine learning?
- Unsupervised Learning vs Supervised vs Decision / reinforcement
- Regression vs Classification
- **Simple Perceptron**
- Neural Networks
- Common Neural Networks
- Activation Functions
- Common Loss functions
- Common Optimizers
- Machine Learning in the context of images
- Ethics



Martin Thoma, CC0, via Wikimedia Commons

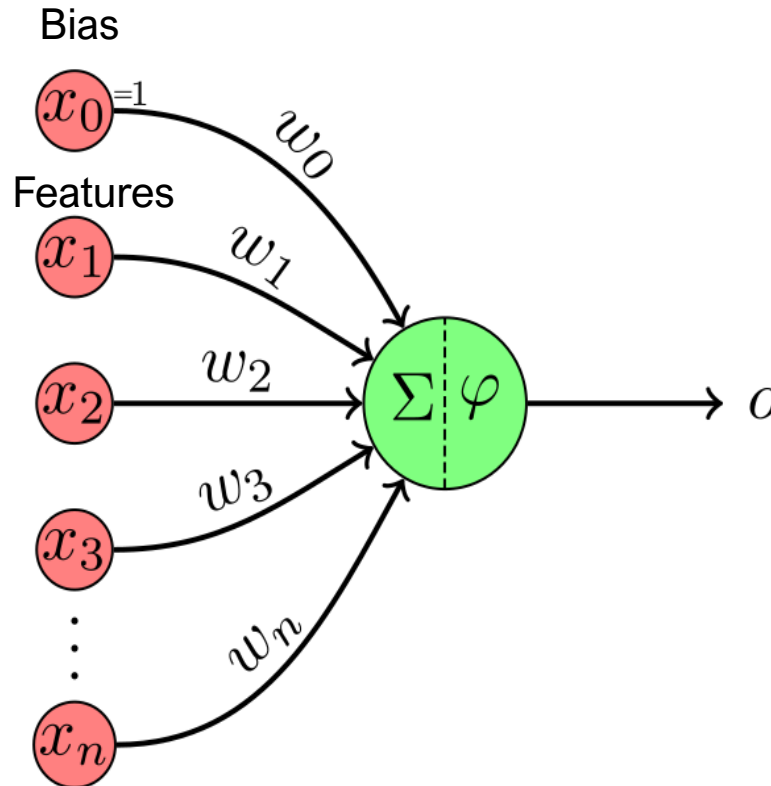
X – data of shape n_features by n_samples
 Y – labels of class -1 or 1 (for case with a decision threshold at 0)



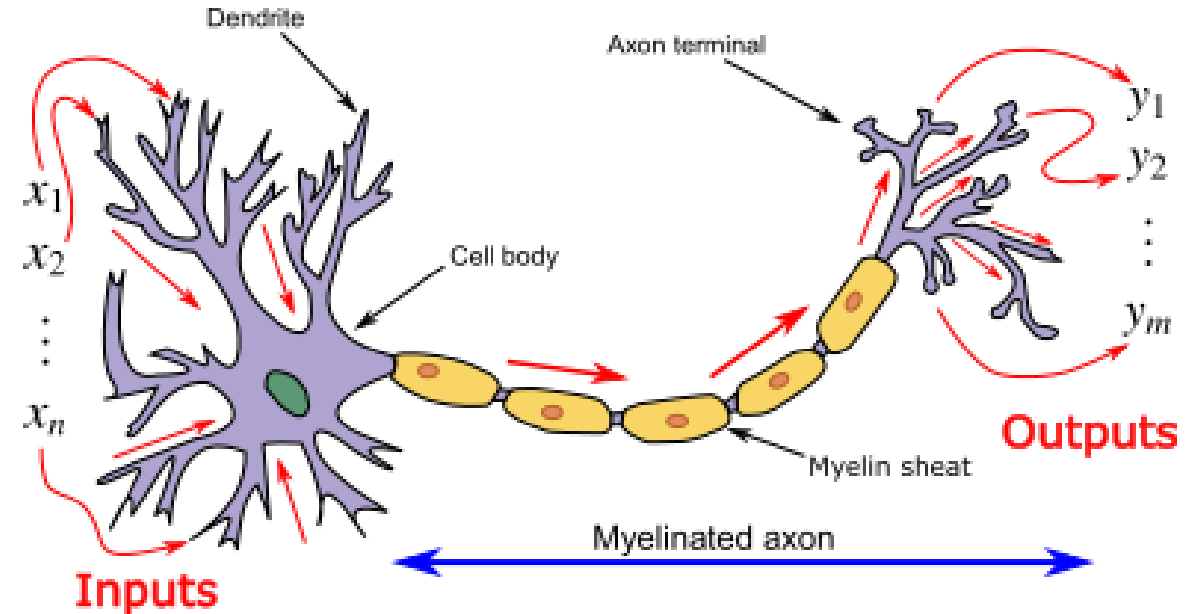
The “Neural” in Neural Nets

Outline

- What is machine learning?
- Unsupervised Learning vs Supervised vs Decision / reinforcement
- Regression vs Classification
- **Simple Perceptron**
- Neural Networks
- Common Neural Networks
- Activation Functions
- Common Loss functions
- Common Optimizers
- Machine Learning in the context of images
- Ethics



Martin Thoma, CC0, via Wikimedia Commons



By Egm4313.s12 (Prof. Loc Vu-Quoc) - Own work, CC BY-SA 4.0, <https://commons.wikimedia.org/w/index.php?curid=72816083>

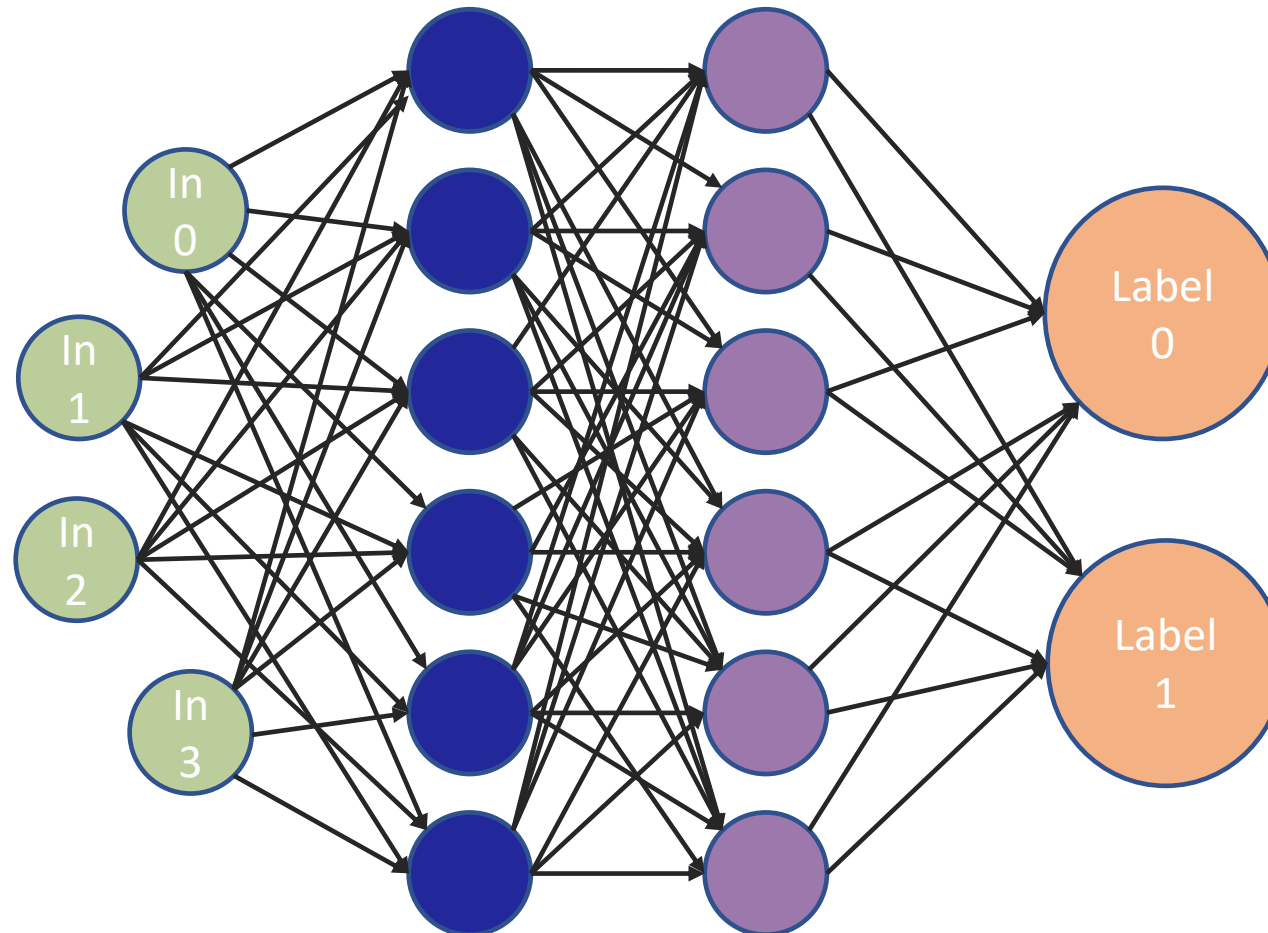
Neural Networks

Outline

- What is machine learning?
- Unsupervised Learning vs Supervised vs Decision / reinforcement
- Regression vs Classification
- Simple Perceptron
- **Neural Networks**
- Common Neural Networks
- Activation Functions
- Common Loss functions
- Common Optimizers
- Machine Learning in the context of images
- Ethics

Feed Forward Neural Networks

What if one daisy chains a ton of perceptrons together?



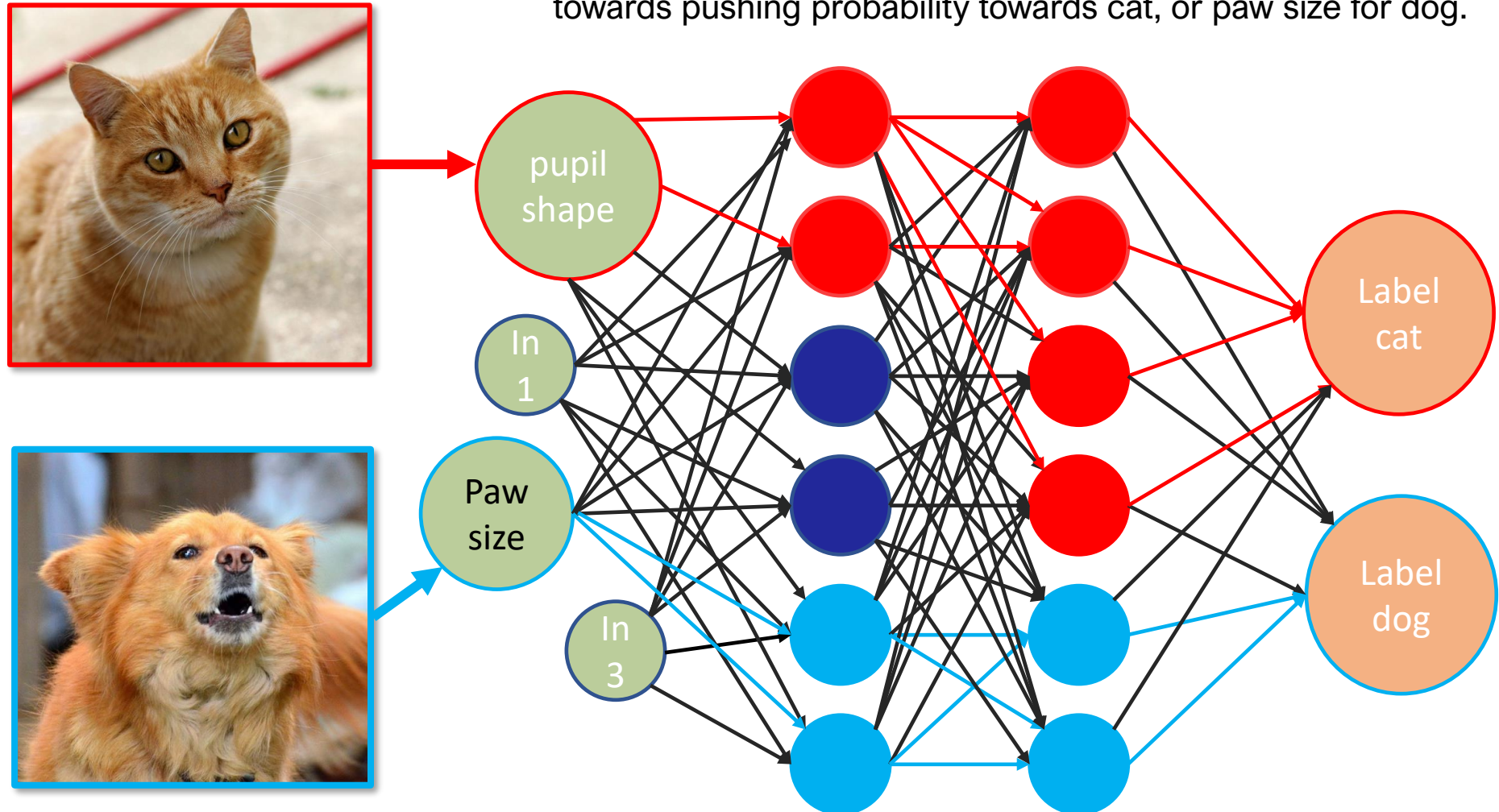
Neural Networks

Outline

- What is machine learning?
- Unsupervised Learning vs Supervised vs Decision / reinforcement
- Regression vs Classification
- Simple Perceptron
- **Neural Networks**
- Common Neural Networks
- Activation Functions
- Common Loss functions
- Common Optimizers
- Machine Learning in the context of images
- Ethics

Feed Forward Neural Networks

FFNNs learn neuron activations based on features given, for example a label for pupil shape may activate neurons specific towards pushing probability towards cat, or paw size for dog.



Note, it's more complicated, but this is a simple conceptual visualization of their function. Features hit activations with different weights that overall lead to a learned decision

Convolutional Neural Networks

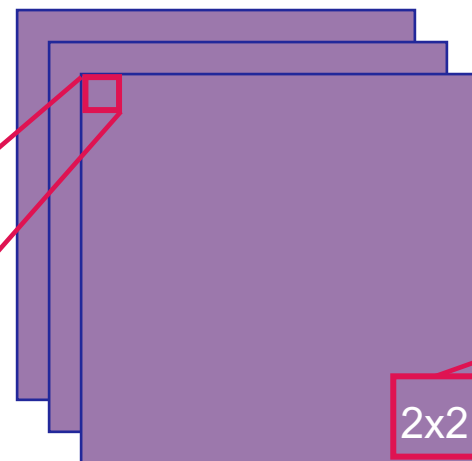
Outline

- What is machine learning?
- Unsupervised Learning vs Supervised vs Decision / reinforcement
- Regression vs Classification
- Simple Perceptron
- Neural Networks
- **Common Neural Networks**
- Activation Functions
- Common Loss functions
- Common Optimizers
- Machine Learning in the context of images
- Ethics

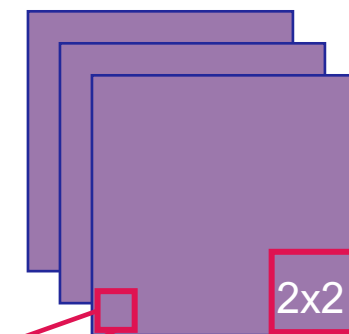
CNN



128x128x3



64x64x64

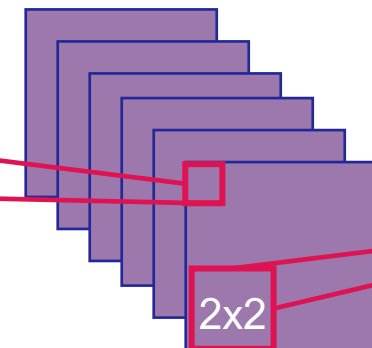


32x32x64

Max pool

$(\text{Input} - \text{kernel window}) / \text{strides} \times$
 $(\text{Input} - \text{kernel window}) / \text{strides} \times$
 Filters

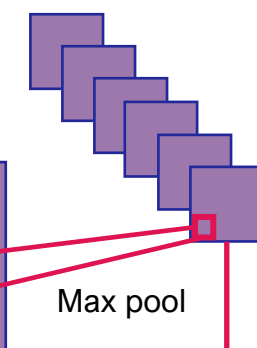
16x16x32



2x2

Max pool

8x8x32

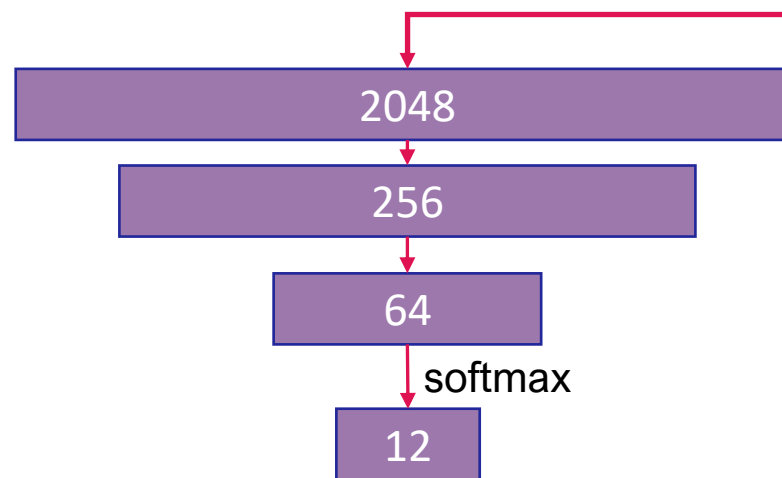


Filters = 64
 kernel = 3x3
 strides = (2,2)

Layer 1: $(128-3)/2 + 1 = 64$

Filters = 32
 kernel = 2x2
 strides = (2,2)

Layer 2: $(32-2)/2 + 1 = 16$

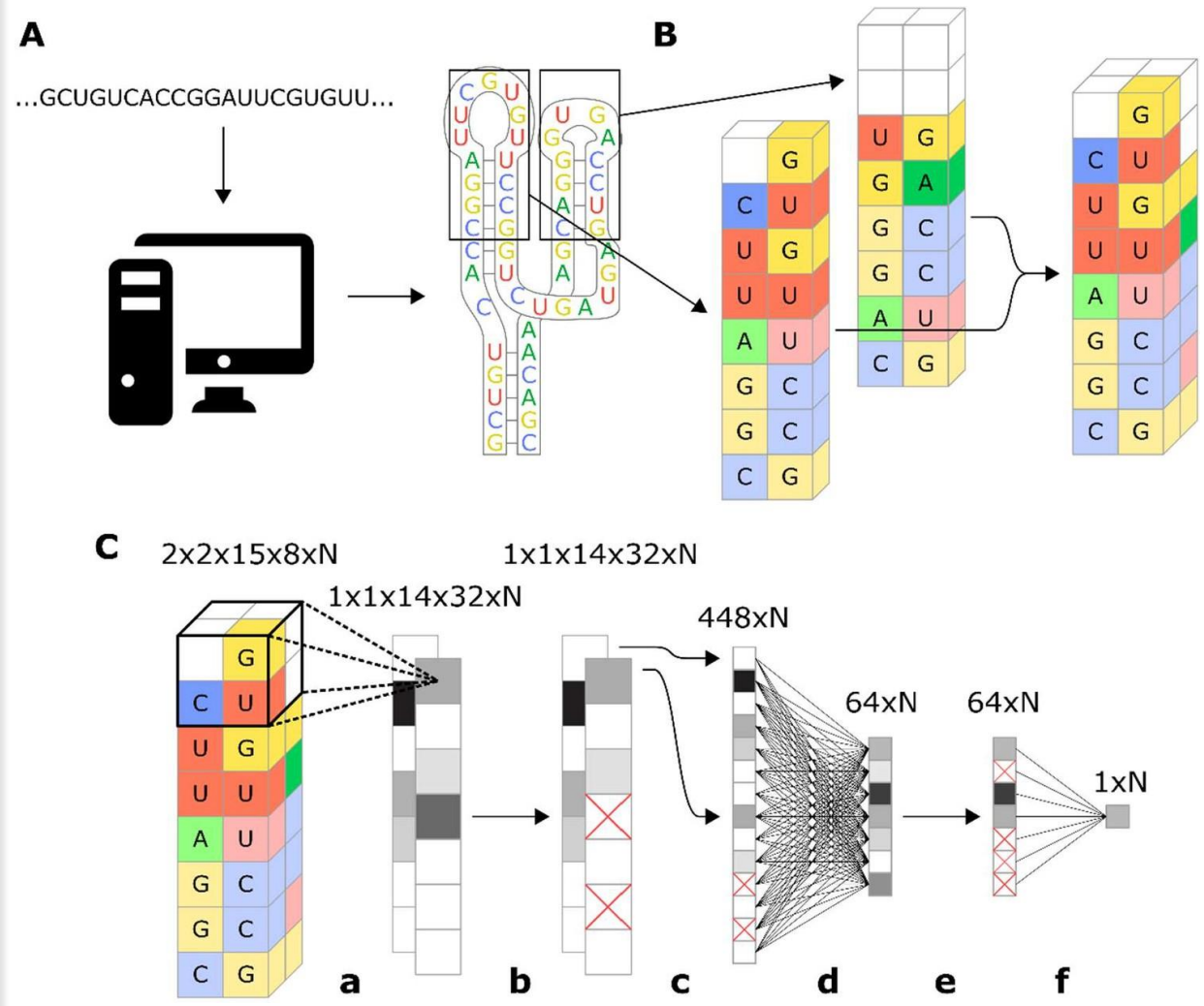


Probability label 0, Probability label 1 ..., Probability label N

Label	Percent
sunflower:	100.0000
buttercup:	0.0000
dandelion:	0.0000
colts foot:	0.0000
daffodil:	0.0000
tulip:	0.0000
tiger lily:	0.0000
crocus:	0.0000
daisy:	0.0000
windflower:	0.0000
cowslip:	0.0000
snowdrop:	0.0000
fritillary:	0.0000
lily valley:	0.0000
iris:	0.0000
bluebell:	0.0000
pansy:	0.0000

CNN - Biological Contexts

- Outline**
- What is machine learning?
 - **Unsupervised Learning vs Supervised vs Decision / reinforcement**
 - Regression vs Classification
 - Simple Perceptron
 - Neural Networks
 - Common Neural Networks
 - Activation Functions
 - Common Loss functions
 - Common Optimizers
 - Machine Learning in the context of images
 - Ethics



A convolutional neural network for the prediction and forward design of ribozyme-based gene-control elements. C Schmidt and C Smolke. eLife 2021. DOI: 10.7554/eLife.59697

Goal: Predict and design RNA Ribozyme Elements

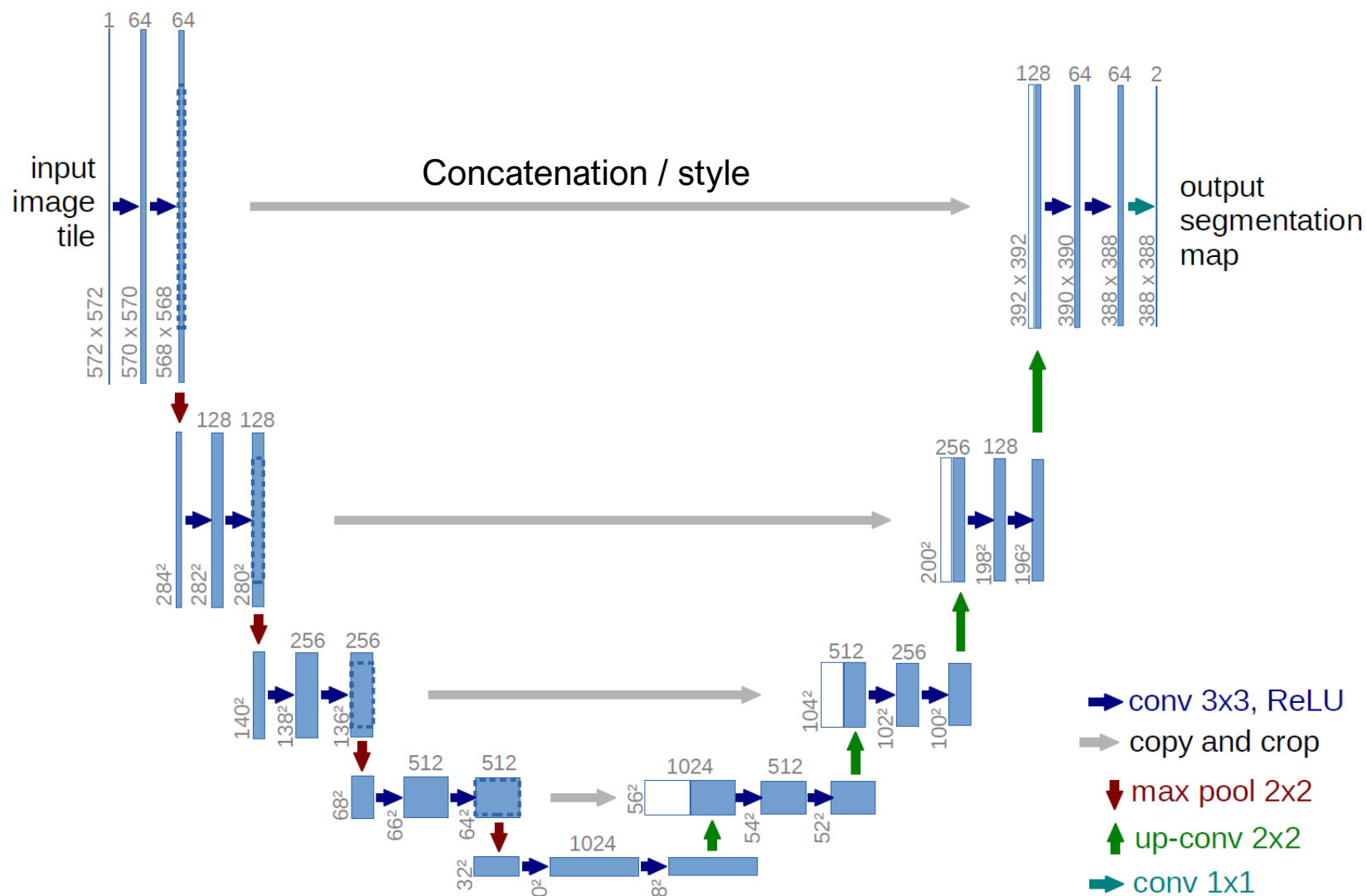
Training data:
 Computational folded Sequence then labeled with loop and branch sizes into a secondary structure feature set

U-Net (Segmentation CNN)

Outline

- What is machine learning?
- Unsupervised Learning vs Supervised vs Decision / reinforcement
- Regression vs Classification
- Simple Perceptron
- Neural Networks
- **Common Neural Networks**
- Activation Functions
- Common Loss functions
- Common Optimizers
- Machine Learning in the context of images
- Ethics

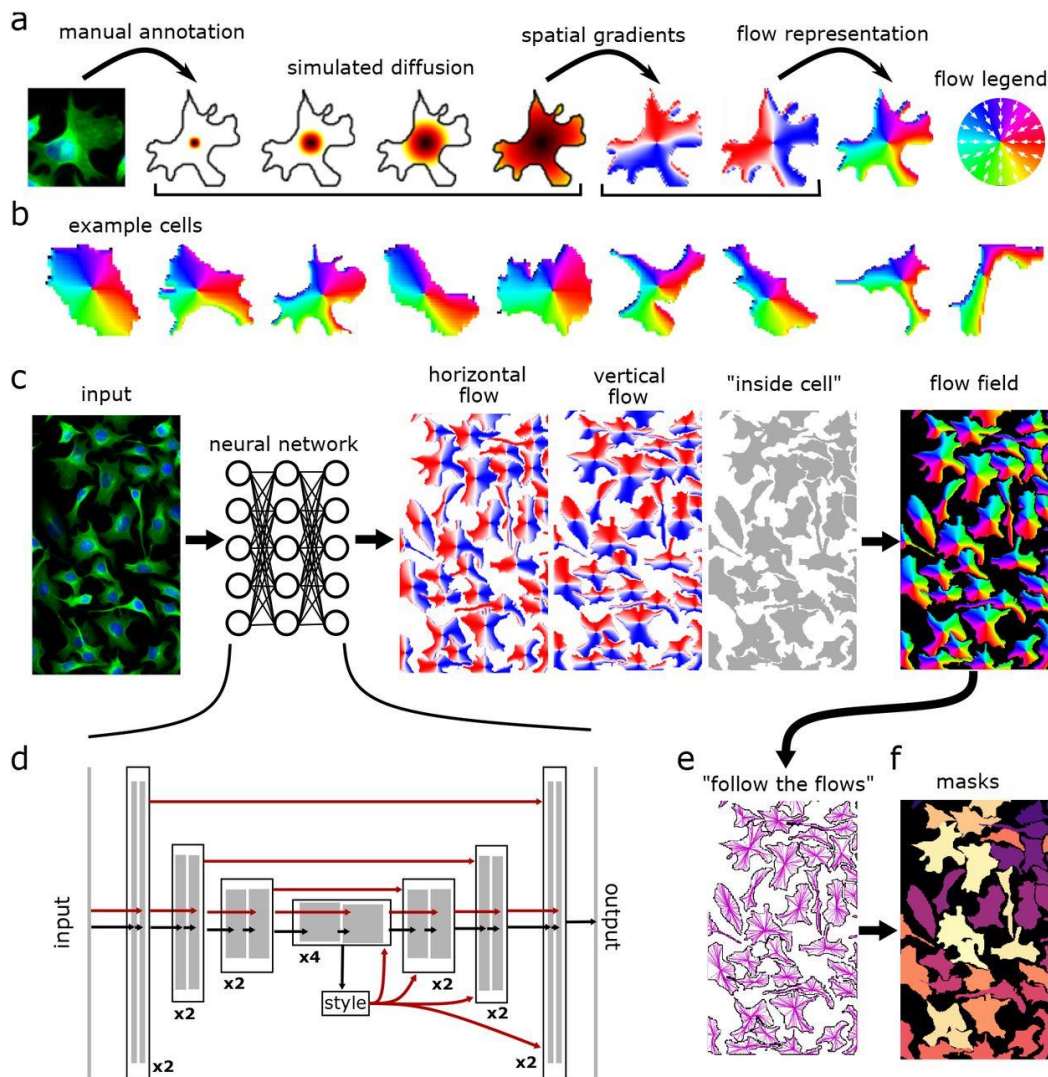
Specific CNN / encoder with a style concatenation of the original feature map in the expansion step. Original implemented for segmenting biomedical images.



U-Net - Biological Contexts

Outline

- What is machine learning?
- **Unsupervised Learning vs Supervised vs Decision / reinforcement**
- Regression vs Classification
- Simple Perceptron
- Neural Networks
- Common Neural Networks
- Activation Functions
- Common Loss functions
- Common Optimizers
- Machine Learning in the context of images
- Ethics



Cellpose: a generalist algorithm for cellular segmentation. Stringer (et al) Nature Methods 2020. DOI: 10.1038/s41592-020-01018-x

Cell Pose

Architecture: Extension of the U-Net with gradient flow

Goal: produce masks of cell images

Training data: hand labeled masks + diffusion simulations

Activation Functions

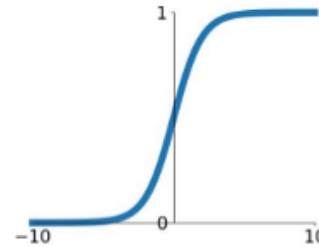
Outline

- What is machine learning?
- Unsupervised Learning vs Supervised vs Decision / reinforcement
- Regression vs Classification
- Simple Perceptron
- Neural Networks
- Common Neural Networks
- **Activation Functions**
- Common Loss functions
- Common Optimizers
- Machine Learning in the context of images
- Ethics

Activation Functions

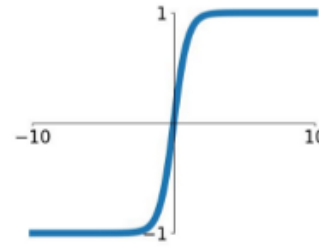
Sigmoid

$$\sigma(x) = \frac{1}{1+e^{-x}}$$



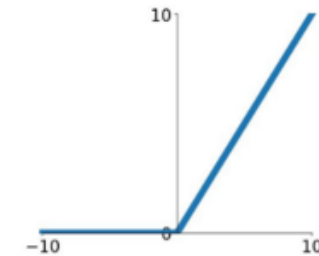
tanh

$$\tanh(x)$$



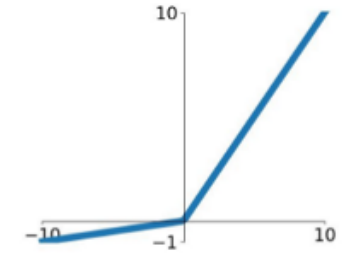
ReLU

$$\max(0, x)$$



Leaky ReLU

$$\max(0.1x, x)$$

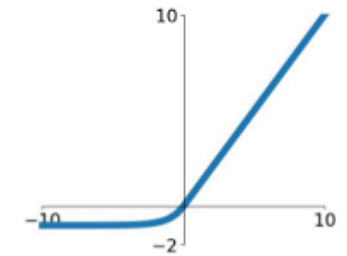


Maxout

$$\max(w_1^T x + b_1, w_2^T x + b_2)$$

ELU

$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$



Source: Shruti Jadon - <https://medium.com/@shrutijadon10104776/survey-on-activation-functions-for-deep-learning-9689331ba092>

Activation Functions

Outline

- What is machine learning?
- Unsupervised Learning vs Supervised vs Decision / reinforcement
- Regression vs Classification
- Simple Perceptron
- Neural Networks
- Common Neural Networks
- **Activation Functions**
- Common Loss functions
- Common Optimizers
- Machine Learning in the context of images
- Ethics

ReLU

- Extreme speed ups of stochastic descent optimizers
- Computationally simple
- Can fail with large gradients – “Dying ReLU”

Leaky ReLU

- Fixes Dying ReLU*
- Computationally simple, but more expensive than ReLU

Random ReLU

- Leaky ReLU with random alphas

Tanh

- Like the Sigmoid, but centered at 0 with -1 and 1
- Better than the Sigmoid in almost every instance

Sigmoid

- Expensive
- Can flatten a gradient, making all outputs zero or one
- Analogous to neuron firing
- Almost never used anymore

Activation functions are application dependent; Take care to either do a literature search for which activation functions are used and why -- or spend time to test different activation function performances on your problem!

<https://stats.stackexchange.com/questions/115258/comprehensive-list-of-activation-functions-in-neural-networks-with-pros-cons>

Loss Functions

Outline

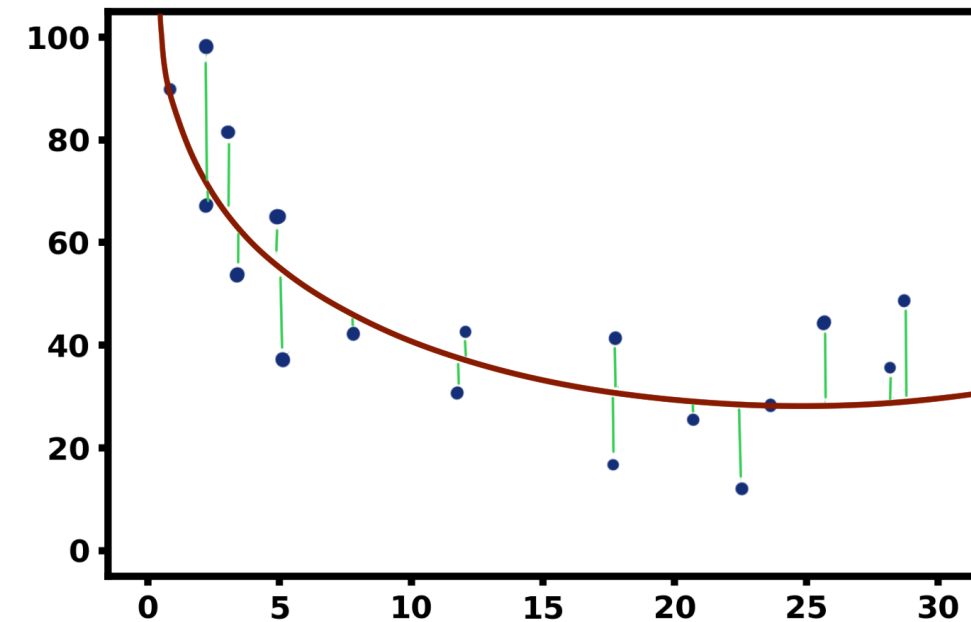
- What is machine learning?
- Unsupervised Learning vs Supervised vs Decision / reinforcement
- Regression vs Classification
- Simple Perceptron
- Neural Networks
- Common Neural Networks
- Activation Functions
- **Common Loss functions**
- Common Optimizers
- Machine Learning in the context of images
- Ethics

How well is the machine learning? Lower the loss, the better!

Mean Squared Error

- The classic!

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$



Log Likelihood

* Probability of data given the model

Cross Entropy Loss

- Binary Classification
- Simplifies from Log Likelihood to an easy expression in the case of labels 0 and 1

$$H(p, q) = - \sum_i p_i \log q_i = -y \log \hat{y} - (1-y) \log(1-\hat{y})$$

p – probability of labels / True labels

q – probability of prediction / Predicted labels

Loss Functions

Outline

- What is machine learning?
- Unsupervised Learning vs Supervised vs Decision / reinforcement
- Regression vs Classification
- Simple Perceptron
- Neural Networks
- Common Neural Networks
- Activation Functions
- **Common Loss functions**
- Common Optimizers
- Machine Learning in the context of images
- Ethics

Hinge loss

- Multiclass classification
- Best for SVM
- Best for labels -1 to 1

Categorical Cross Entropy

- Multiclass classification

Sparse Categorical Cross Entropy

- Multiclass classification
- One hot encoding on labels

Kullback Leibler Divergence

- Multiclass classification
- One hot encoding on labels

And many more...

Functions

`KLD(...)`: Computes Kullback-Leibler divergence loss between `y_true` and `y_pred`.

`MAE(...)`: Computes the mean absolute error between labels and predictions.

`MAPE(...)`: Computes the mean absolute percentage error between `y_true` and `y_pred`.

`MSE(...)`: Computes the mean squared error between labels and predictions.

`MSLE(...)`: Computes the mean squared logarithmic error between `y_true` and `y_pred`.

`binary_crossentropy(...)`: Computes the binary crossentropy loss.

`categorical_crossentropy(...)`: Computes the categorical crossentropy loss.

`categorical_hinge(...)`: Computes the categorical hinge loss between `y_true` and `y_pred`.

`cosine_similarity(...)`: Computes the cosine similarity between labels and predictions.

`deserialize(...)`: Deserializes a serialized loss class/function instance.

`get(...)`: Retrieves a Keras loss as a `function / Loss` class instance.

`hinge(...)`: Computes the hinge loss between `y_true` and `y_pred`.

`huber(...)`: Computes Huber loss value.

`kl_divergence(...)`: Computes Kullback-Leibler divergence loss between `y_true` and `y_pred`.

`klD(...)`: Computes Kullback-Leibler divergence loss between `y_true` and `y_pred`.

`kullback_leibler_divergence(...)`: Computes Kullback-Leibler divergence loss between `y_true` and `y_pred`.

`log_cosh(...)`: Logarithm of the hyperbolic cosine of the prediction error.

`logcosh(...)`: Logarithm of the hyperbolic cosine of the prediction error.

`mae(...)`: Computes the mean absolute error between labels and predictions.

`mape(...)`: Computes the mean absolute percentage error between `y_true` and `y_pred`.

https://www.tensorflow.org/api_docs/python/tf/keras/losses

Loss Functions

Outline

- What is machine learning?
- Unsupervised Learning vs Supervised vs Decision / reinforcement
- Regression vs Classification
- Simple Perceptron
- Neural Networks
- Common Neural Networks
- Activation Functions
- **Common Loss functions**
- Common Optimizers
- Machine Learning in the context of images
- Ethics

Functions

- `KLD(...)`: Computes Kullback-Leibler divergence loss between `y_true` and `y_pred`.
- `MAE(...)`: Computes the mean absolute error between labels and predictions.
- `MAPE(...)`: Computes the mean absolute percentage error between `y_true` and `y_pred`.
- `MSE(...)`: Computes the mean squared error between labels and predictions.
- `MSLE(...)`: Computes the mean squared logarithmic error between `y_true` and `y_pred`.
- `binary_crossentropy(...)`: Computes the binary crossentropy loss.
- `categorical_crossentropy(...)`: Computes the categorical crossentropy loss.
- `categorical_hinge(...)`: Computes the categorical hinge loss between `y_true` and `y_pred`.
- `cosine_similarity(...)`: Computes the cosine similarity between labels and predictions.
- `deserialize(...)`: Deserializes a serialized loss class/function instance.
- `get(...)`: Retrieves a Keras loss as a `function / Loss` class instance.
- `hinge(...)`: Computes the hinge loss between `y_true` and `y_pred`.
- `huber(...)`: Computes Huber loss value.
- `kl_divergence(...)`: Computes Kullback-Leibler divergence loss between `y_true` and `y_pred`.
- `kld(...)`: Computes Kullback-Leibler divergence loss between `y_true` and `y_pred`.
- `kullback_leibler_divergence(...)`: Computes Kullback-Leibler divergence loss between `y_true` and `y_pred`.
- `log_cosh(...)`: Logarithm of the hyperbolic cosine of the prediction error.
- `logcosh(...)`: Logarithm of the hyperbolic cosine of the prediction error.
- `mae(...)`: Computes the mean absolute error between labels and predictions.
- `mape(...)`: Computes the mean absolute percentage error between `y_true` and `y_pred`.

https://www.tensorflow.org/api_docs/python/tf/keras/losses

Once again, loss functions are can get very granular for your specific application, once again highlighting the need for a strong literature search when setting out to generate new models. Additionally, you may take a programmatic search and test multiple loss functions for your application, or even use a sum of loss functions to train.

Outline

- What is machine learning?
- Unsupervised Learning vs Supervised vs Decision / reinforcement
- Regression vs Classification
- Simple Perceptron
- Neural Networks
- Common Neural Networks
- Activation Functions
- Common Loss functions
- **Common Optimizers**
- Machine Learning in the context of images
- Ethics

CONSIDERATIONS:

- Propensity for local minima sticking (globalness?)
- Computational Cost
- Dynamic or Static Learning Rates
- Memory / Batch sizes

Stochastic Gradient Descent

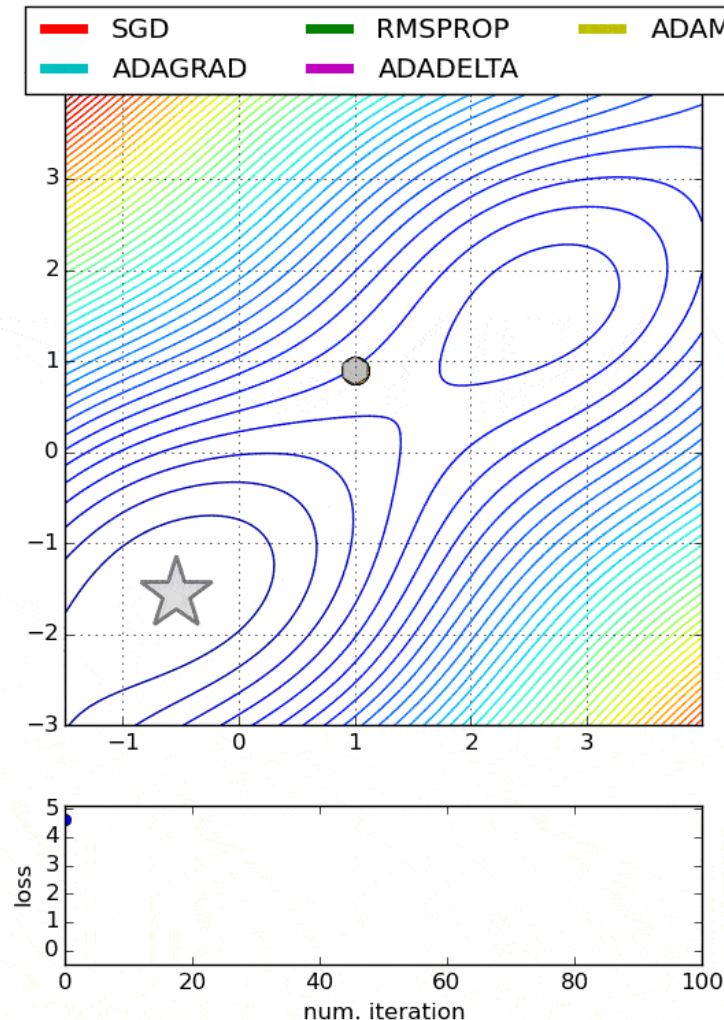
- Stochastically move towards the minimum based on each new data point

RMSprop

- Adaptive learning rate
- RMS average of the squared gradients for each weight

Adam

- Adaptive moments that keeps track of momentum (1st and 2nd moments) and corrects based on a learned decay rate



<https://towardsdatascience.com/optimizers-for-training-neural-network-59450d71caf6>

https://moodle2.cs.huji.ac.il/nu15/pluginfile.php/316969/mod_resource/content/1/adam_pres.pdf

Evaluating Models

Outline

- What is machine learning?
- Unsupervised Learning vs Supervised vs Decision / reinforcement
- Regression vs Classification
- Simple Perceptron
- Neural Networks
- Common Neural Networks
- Activation Functions
- Common Loss functions
- Common Optimizers
- **Evaluating Models**
- Ethics

100 Cat photos



100 Dog photos



3 cats mislabeled as dogs
17 dogs mislabeled as cats



Confusion Matrix

Accuracy: $(97 + 83) / (100 + 100)$

	Predicted Cat	Predicted Dog
Actual Cat	97	3
Actual Dog	17	83

Evaluating Models

Outline

- What is machine learning?
- Unsupervised Learning vs Supervised vs Decision / reinforcement
- Regression vs Classification
- Simple Perceptron
- Neural Networks
- Common Neural Networks
- Activation Functions
- Common Loss functions
- Common Optimizers
- **Evaluating Models**
- Ethics

Confusion Matrices – Starting with these a plethora of metrics can be calculated for model selection or evaluation

Confusion Matrix

Metrics

		Predicted condition		Sources: [13][14][15][16][17][18][19][20] view • talk • edit	
		Predicted condition positive (PP)	Predicted condition negative (PN)	Informedness, bookmaker informedness (BM) = $TPR + TNR - 1$	Prevalence threshold (PT) = $\frac{\sqrt{TPR \cdot FPR} - FPR}{TPR - FPR}$
Actual condition	Total population = $P + N$				
	Actual condition positive (P)	True positive (TP), hit	False negative (FN), Type II error, miss, overestimation	True positive rate (TPR), recall, sensitivity (SEN), probability of detection, hit rate, power = $\frac{TP}{P}$ = $1 - FNR$	False negative rate (FNR), miss rate = $\frac{FN}{P} = 1 - TPR$
Actual condition negative (N)	False positive (FP), Type I error, false alarm, underestimation	True negative (TN), correct rejection	False positive rate (FPR), probability of false alarm, fall-out = $\frac{FP}{N} = 1 - TNR$	True negative rate (TNR), specificity (SPC), selectivity = $\frac{TN}{N}$ = $1 - FPR$	
	Prevalence = $\frac{P}{P + N}$	Positive predictive value (PPV), precision = $\frac{TP}{PP}$ = $1 - FDR$	False omission rate (FOR) = $\frac{FN}{PN} = 1 - NPV$	Positive likelihood ratio (LR+) = $\frac{TPR}{FPR}$	Negative likelihood ratio (LR-) = $\frac{FNR}{TNR}$
	Accuracy (ACC) = $\frac{TP + TN}{P + N}$	False discovery rate (FDR) = $\frac{FP}{PP} = 1 - PPV$	Negative predictive value (NPV) = $\frac{TN}{PN}$ = $1 - FOR$	Markedness (MK), deltaP (Δp) = $PPV + NPV - 1$	Diagnostic odds ratio (DOR) = $\frac{LR+}{LR-}$
	Balanced accuracy (BA) = $\frac{TPR + TNR}{2}$	F_1 score = $\frac{2 \cdot PPV \cdot TPR}{PPV + TPR} = \frac{2TP}{2TP + FP + FN}$	Fowlkes–Mallows index (FM) = $\sqrt{PPV \cdot TPR}$	Threat score (TS), critical success index (CSI) = $\frac{TP}{TP + FN + FP}$	Matthews correlation coefficient (MCC) = $\frac{\sqrt{TPR \cdot TNR \cdot PPV \cdot NPV} - \sqrt{FNR \cdot FPR \cdot FOR \cdot FDR}}$

https://en.wikipedia.org/wiki/Confusion_matrix

Evaluating Models

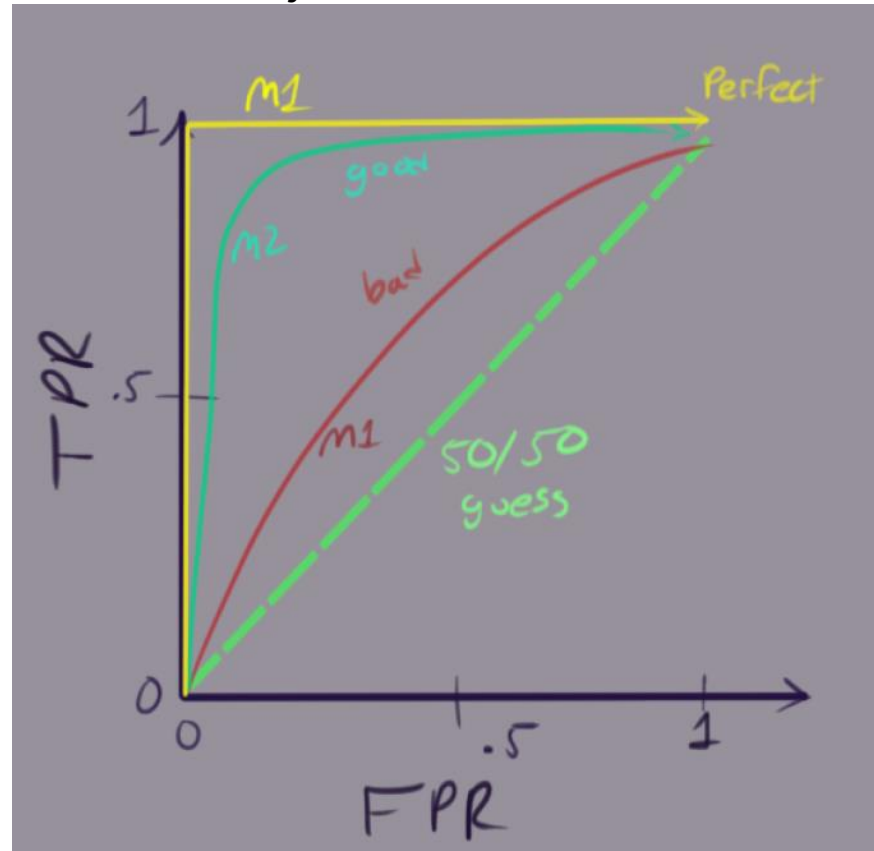
Outline

- What is machine learning?
- Unsupervised Learning vs Supervised vs Decision / reinforcement
- Regression vs Classification
- Simple Perceptron
- Neural Networks
- Common Neural Networks
- Activation Functions
- Common Loss functions
- Common Optimizers
- **Evaluating Models**
- Ethics

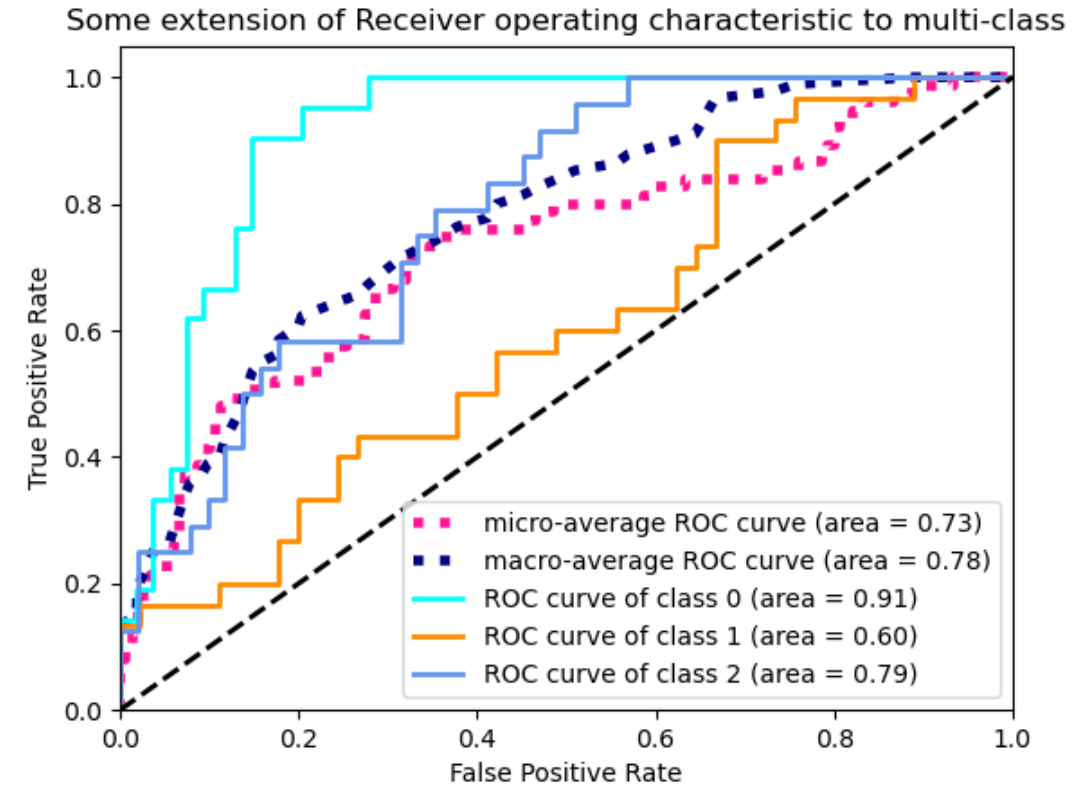
Receiver Operator Characteristic (ROC Curve)

Standard way to compare model performances (accuracy wise, not computational wise)

Binary classification



Mutliclass



https://scikit-learn.org/stable/auto_examples/model_selection/plot_roc.html

Outline

- What is machine learning?
- Unsupervised Learning vs Supervised vs Decision / reinforcement
- Regression vs Classification
- Simple Perceptron
- Neural Networks
- Common Neural Networks
- Activation Functions
- Common Loss functions
- Common Optimizers
- Machine Learning in the context of images
- **Ethics**

Reinforcement of socioeconomic trends

Uncomprehensive datasets

Responsibility and Liability

Lack of transparent internal workings (without much scrutiny)

Consent in dataset building

Resources:

<https://www.nature.com/articles/s41599-020-0501-9>

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5962261/>

http://www3.weforum.org/docs/WEF_40065_White_Paper_How_to_Prevent_Discriminatory_Outcomes_in_Machine_Learning.pdf