Neural Networks & Deep Learning

Tamara Sipes, PhD

Director of Commercial Data Science, OptumInsight, Payer Analytics

OPTUM Labs®

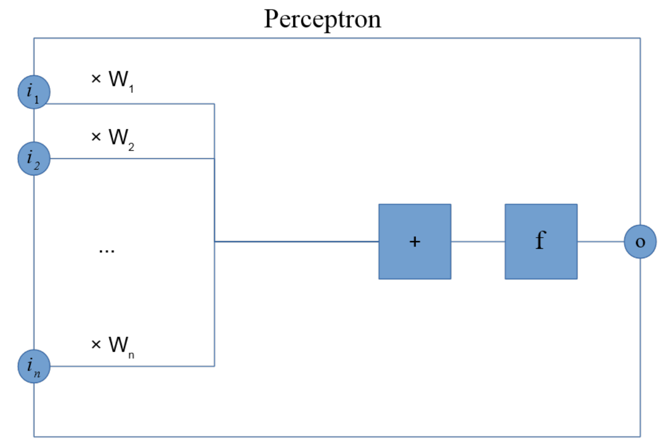
Disclosures

Financial disclosures:

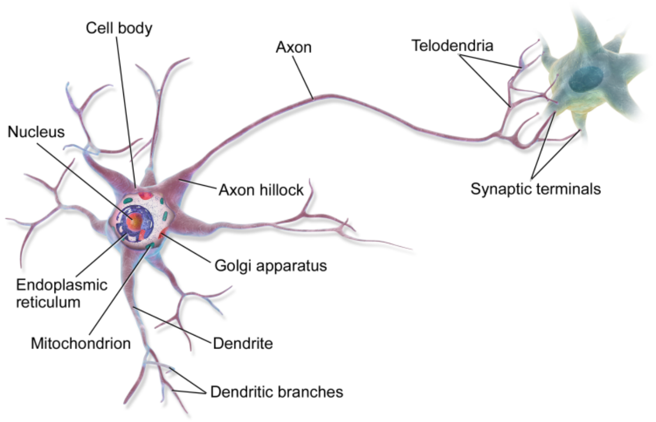
- CureMetrix, Inc., co-founder with founder shares

Perceptron: An “artificial neuron”

Perceptron & biological neuron



$$o = f\left(\sum_{k=1}^{n} i_k \cdot W_k\right)$$



- Inspired by biological neural networks. Perceptron has several inputs.
- Multiple inputs are multiplied by weights assigned to that input.
- An offset (bias) is added to the weighted sum, which is passed to a function that provides the perceptron output.

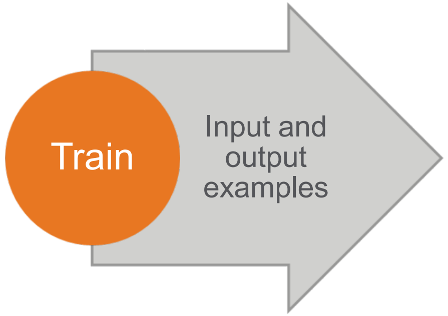
- Dendrites receive signals, cell body processes them, and an exon sends signals to other neurons.
- Multiple inputs are multiplied by weights assigned to that input.
- If the output of a neuron surpasses a certain threshold, the neuron transmits this electrical signal along the axon.

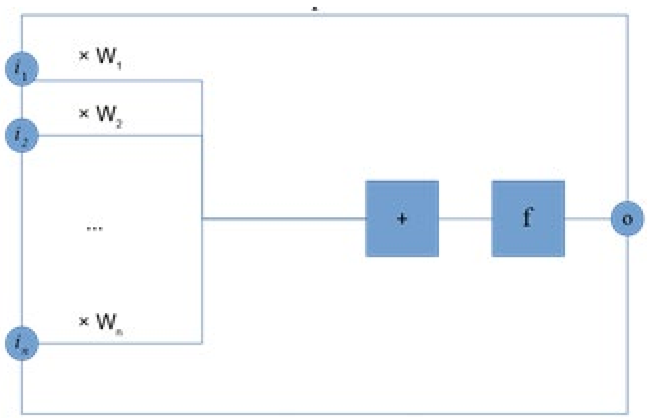
Purves, Dale (2011). Neuroscience (5th ed.). Sunderland, Mass.: Sinauer. p. 507

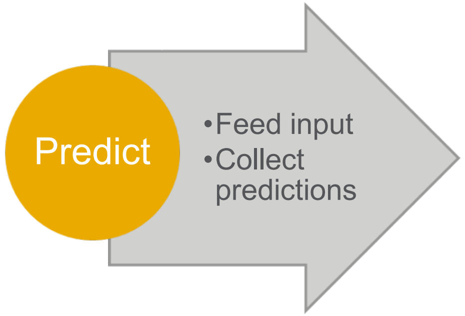
Rosenblatt, Frank (1957), The Perceptron - a perceiving and recognizing automaton. Report 85-460-1, Cornell Aeronautical Laboratory.

Perceptron is a linear classifier

- Perceptron is a supervised learning algorithm (trained on examples of data with known outcomes).
- Once trained, it is used to classify (or, predict) the outcome for new data.



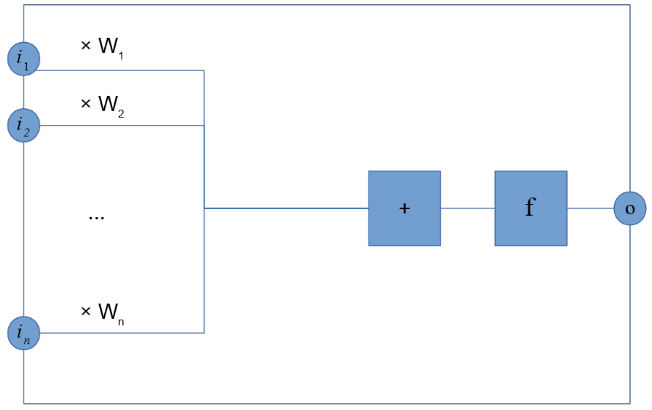




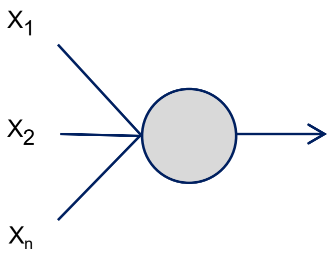
- Perceptron is a linear classifier – it uses a linear prediction function (a sum of the product of weights and inputs).

Perceptron & logistic regression

Perceptron



$$o = f\left(\sum_{k=1}^{n} i_k \cdot W_k\right)$$

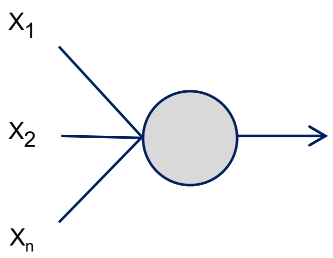


$$\mathrm{h}_\theta(x) = \frac{1}{1 + \exp(-\theta^T x)}$$

- Logistic regression is similar in structure, having one unit and multiple inputs, but logistic regression has probabilistic connotations.
- LR provides a measure of uncertainty in the occurrence of a binary outcome (0 or 1). The output is bounded asymptotically between 0 and 1.
- Both can be used as building blocks to create more complex classifiers.

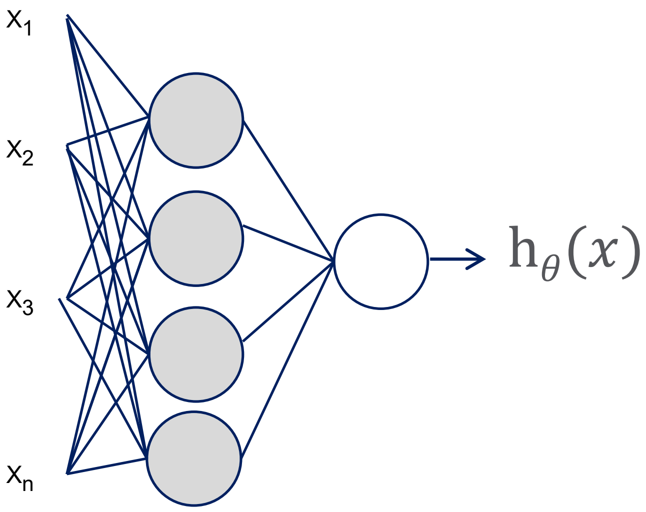
Logistic regression vs. artificial neural network (ANN)

LR:

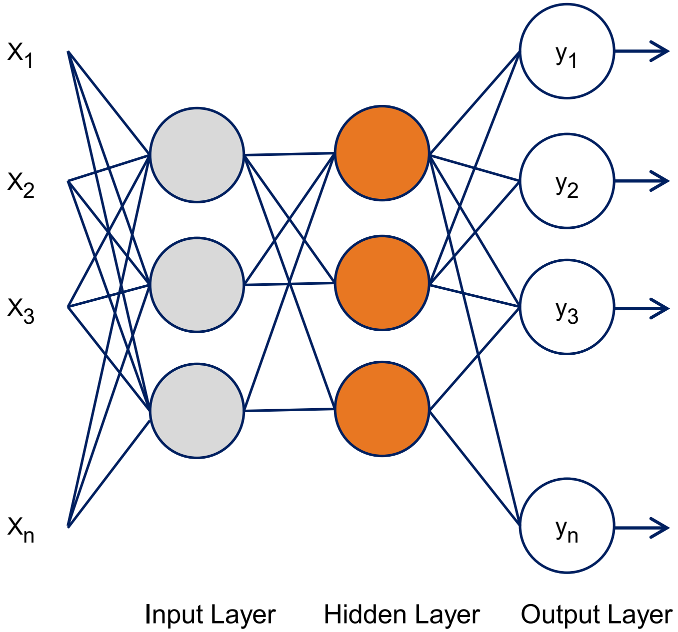


$$h_\theta(x) = \frac{1}{1+\exp(-\theta^T x)}$$

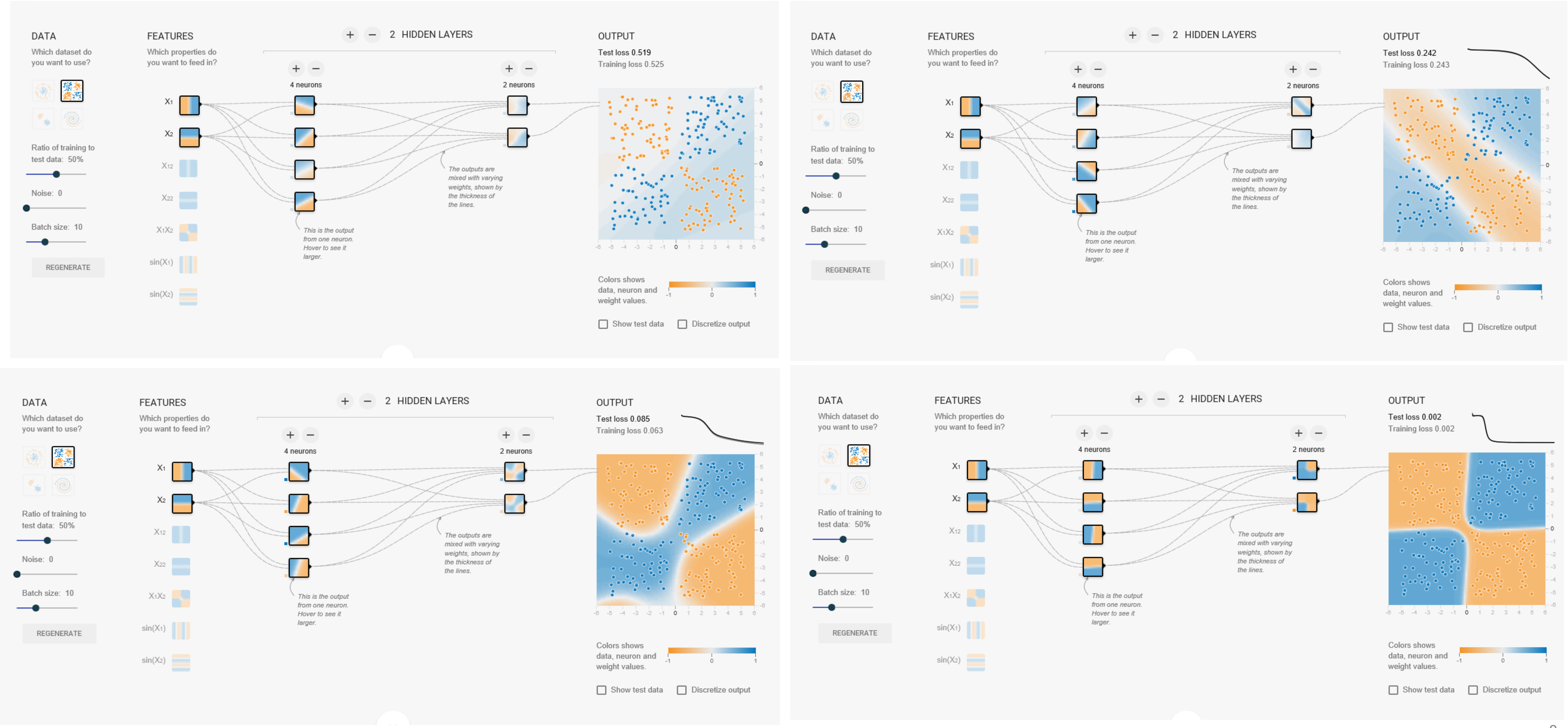
Several LRs strung together = ANN:



ANN with more layers/outputs:

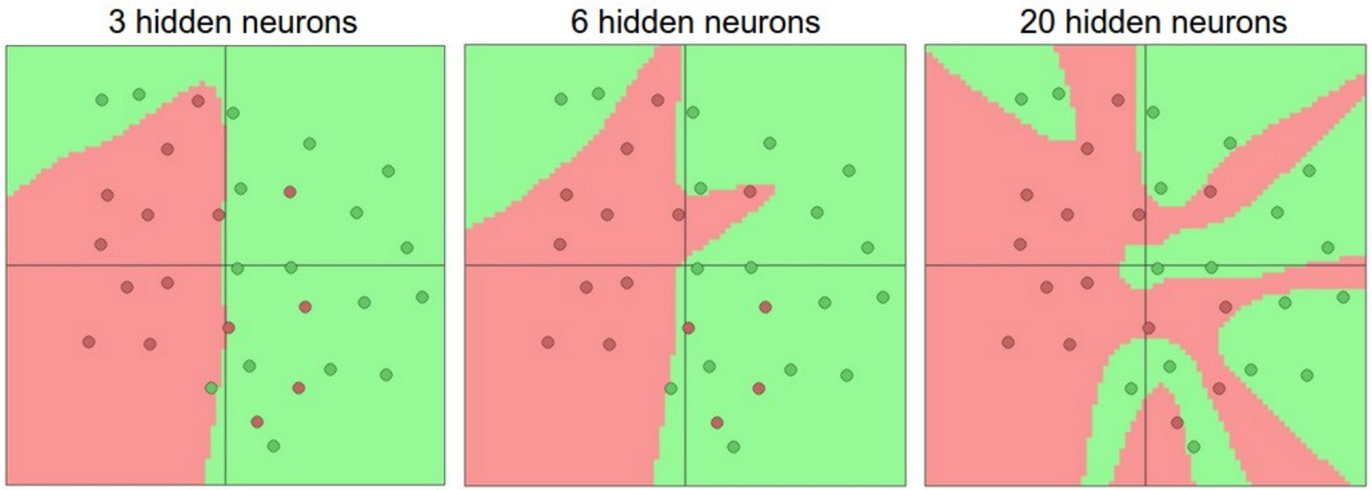


How do neural networks learn?



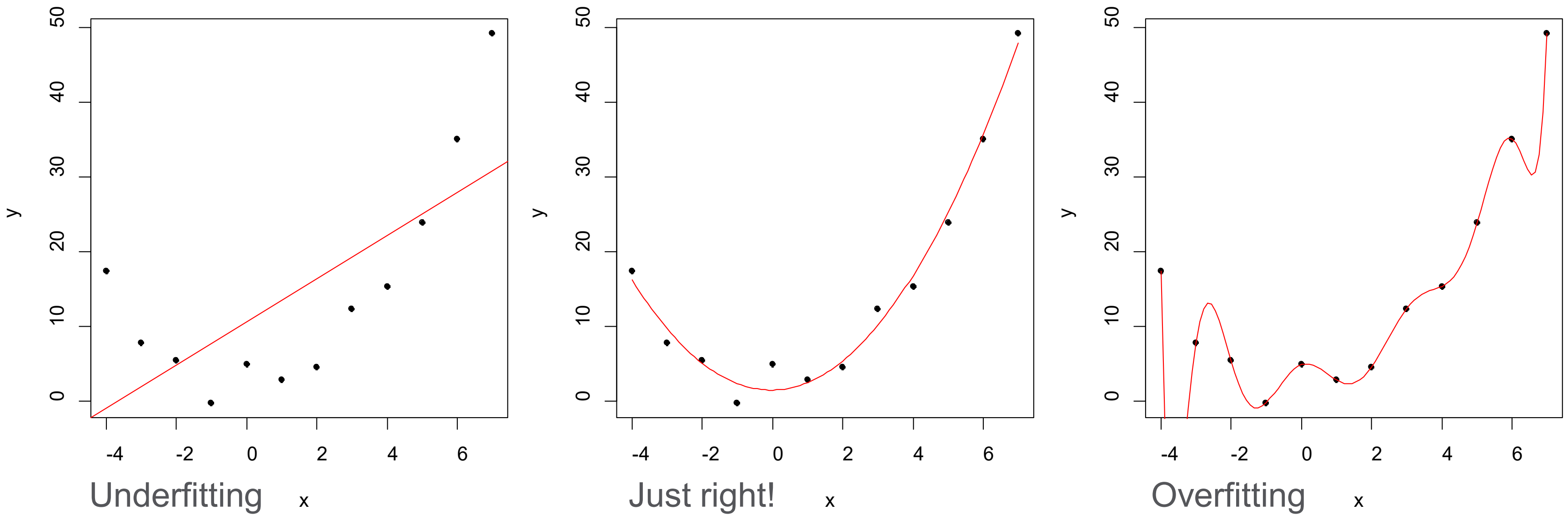
https://playground.tensorflow.org/

Nonlinearity to the rescue: Learning in complex domains



By Andrej Karpathy, http://cs231n.github.io/neural-networks-1/

Tuning ANN: To learn every detail or not?



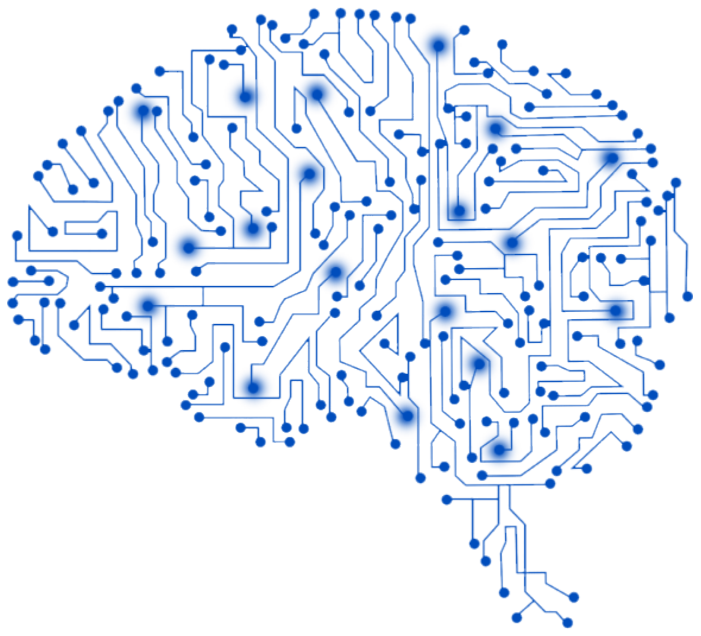
Underfitting Just right! Overfitting

The goal: find the ANN with the best generalization properties and avoid underfitting or overfitting.

Deep learning

What are deep neural networks?

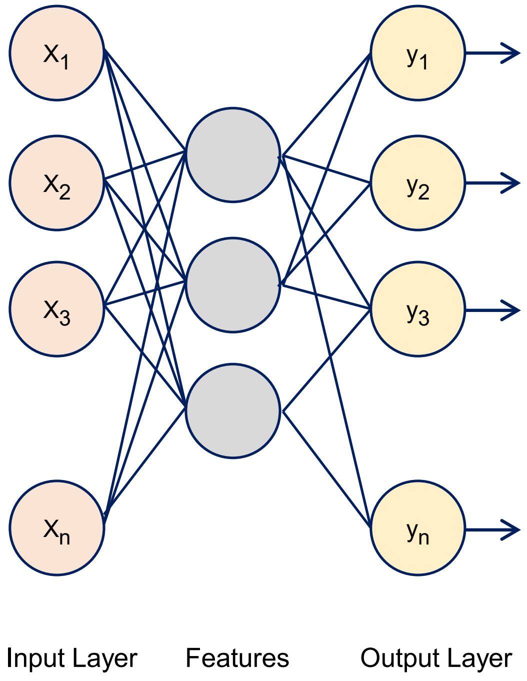
- Deep Neural Networks = neural networks with several hidden layers
- Deep Nets: unsupervised or supervised



https://earthsky.org/upl/2018/05/deep-learning-quora.png

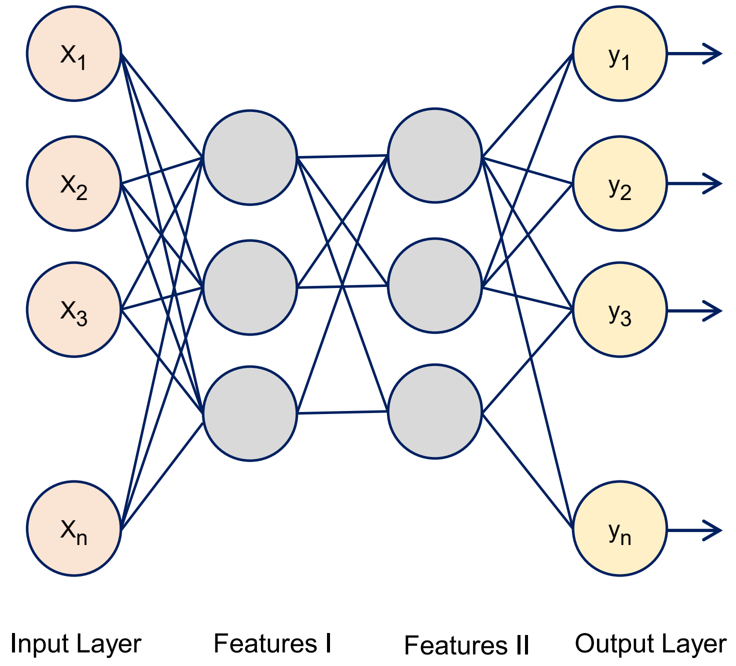
Autoencoder: An introduction

- A type of unsupervised ANN which tries to discover a compressed representation for a dataset
- Architecture:
  - an input layer
  - an output layer of the same size,
  - one or more hidden layers of different sizes connecting input to the output
- An autoencoder is trained to reconstruct its own inputs (hence the *auto*)

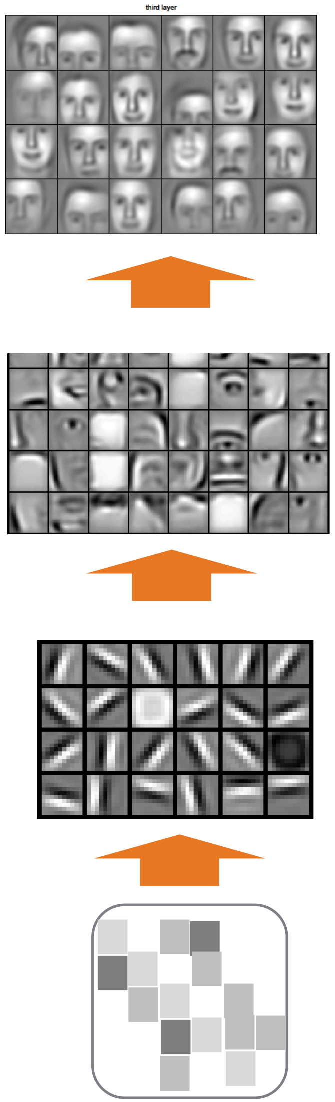


Stacked autoencoders

- Stack several autoencoders and train them using layer-wise training



Feature hierarchies



images

combinations of edges

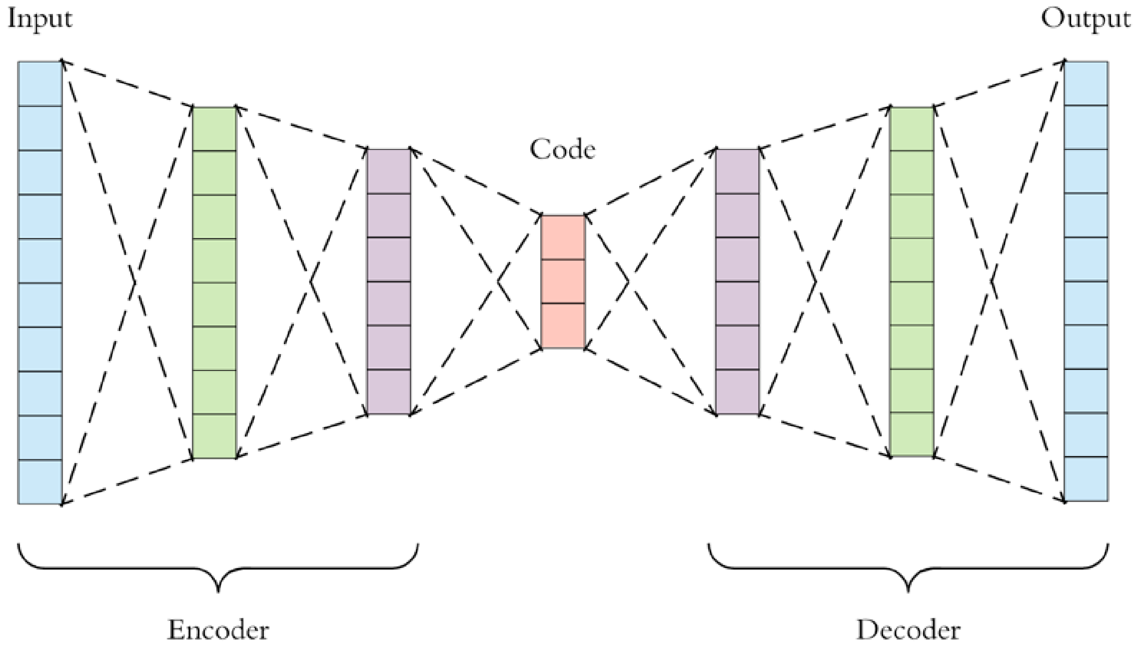
edges

pixels

Adapted from Andrew Ng: “Advanced Topics”

Deep learning based autoencoders

- Create a multi-layer neural network to encode the data into a smaller or simpler representation such that the original data can be reconstructed with high fidelity.



Arden Dertat, Oct 3, 2017. Applied Deep Learning – Part 3: Autoencoders.
https://towardsdatascience.com/applied-deep-learning-part-3-autoencoders-1c083af4d798

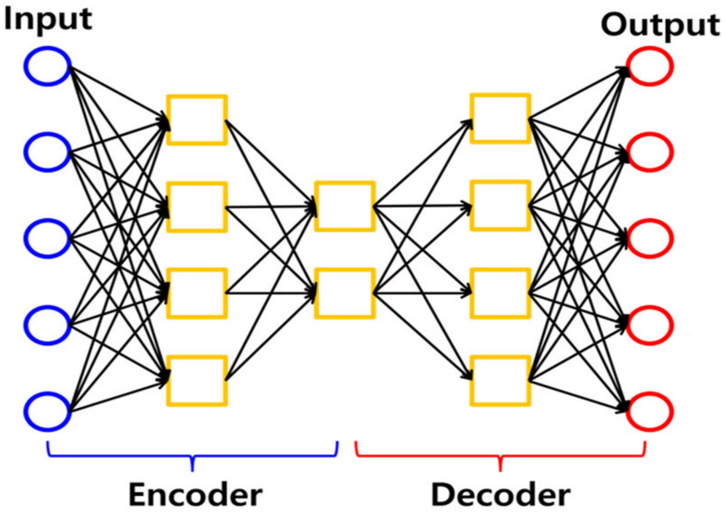
Autoencoder example

- Trained on MNIST data: a “classic” (widely used) ML research

dataset that contains ~10000 images of handwritten digits,

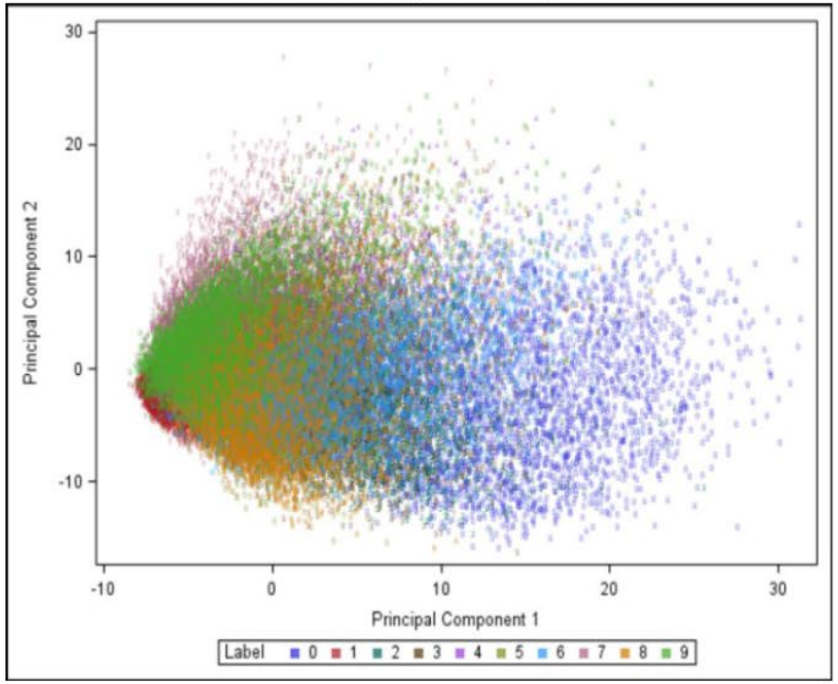
with labels from 0 through 9:

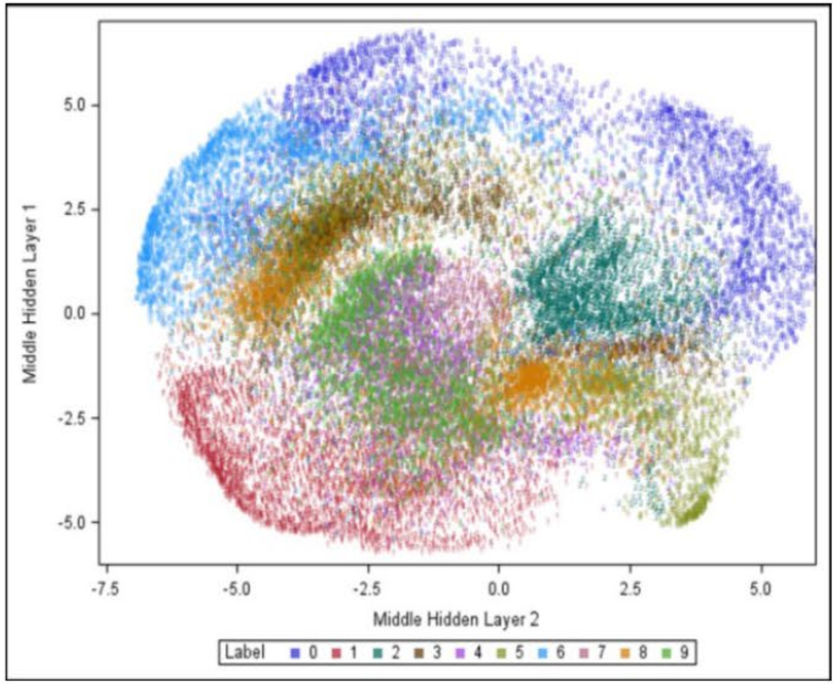
- Hour-glass architecture: 300x100x2x100x300 hidden units:



https://www.researchgate.net/profile/Byoungil-Jeon/publication/341521530/figure/fig1/AS:893277582655489@1589985380689/Schematic-of-autoencoder-architecture.jpg

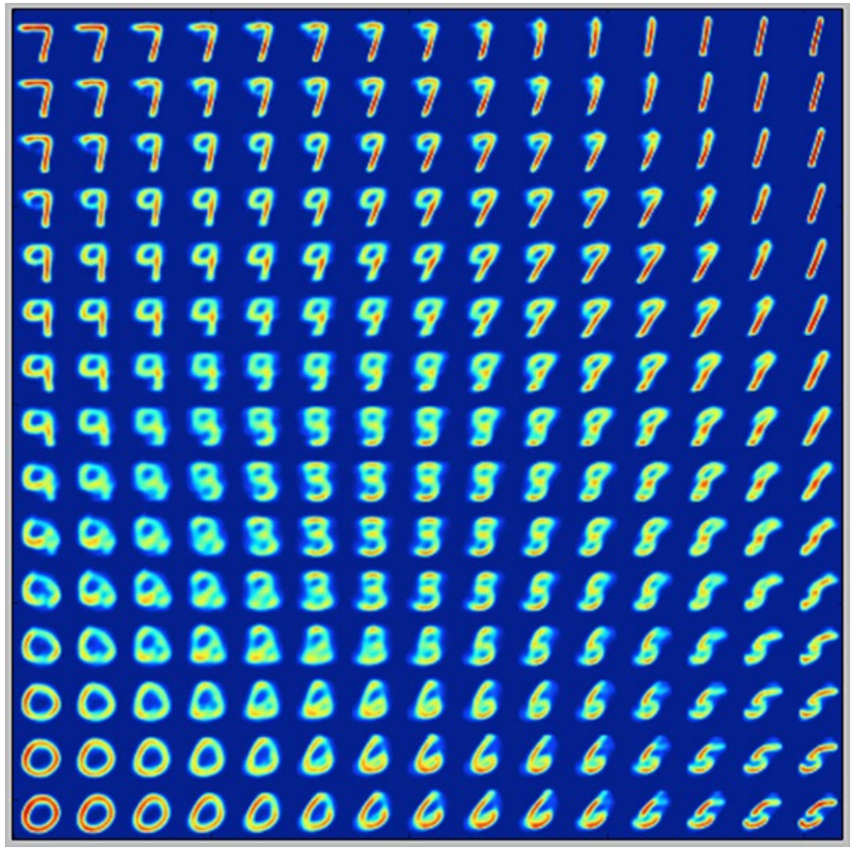
First two principal components vs. middle hidden layer





https://support.sas.com/resources/papers/proceedings14/SAS313-2014.pdf

Computer-generated digits



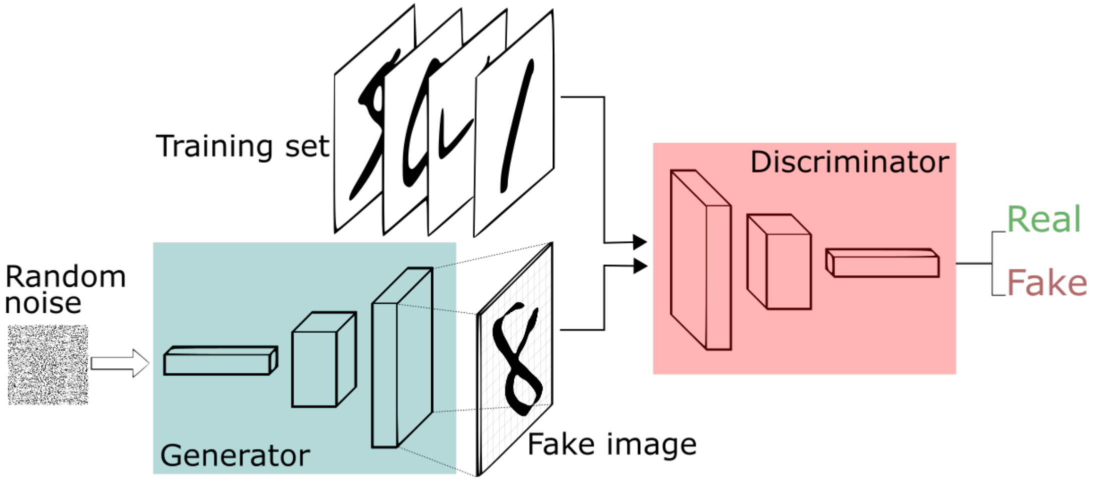
https://blog.keras.io/building-autoencoders-in-keras.html

GANs

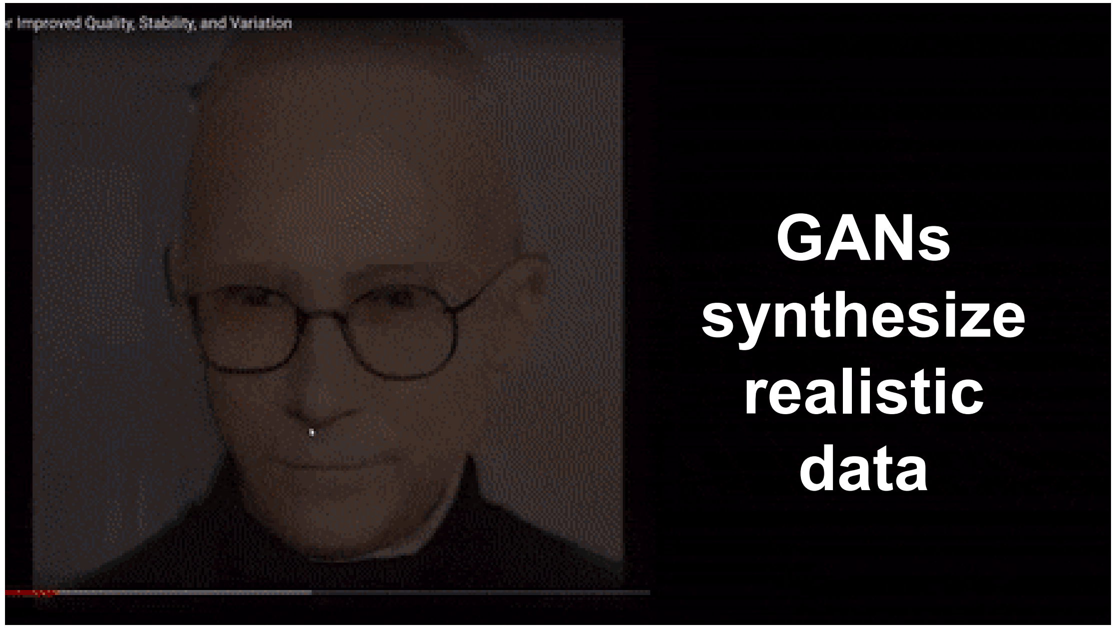
Generative Adversarial Networks

- One model generates examples
- Another model discriminates between those examples and a training set

Improved generation of realistic data



Example: artificial celebrity faces using GANs



https://www.eventbrite.com/e/learning-to-build-generative-adversarial-networks-gans-to-generate-images-tickets-44337796543
https://arxiv.org/abs/1812.04948

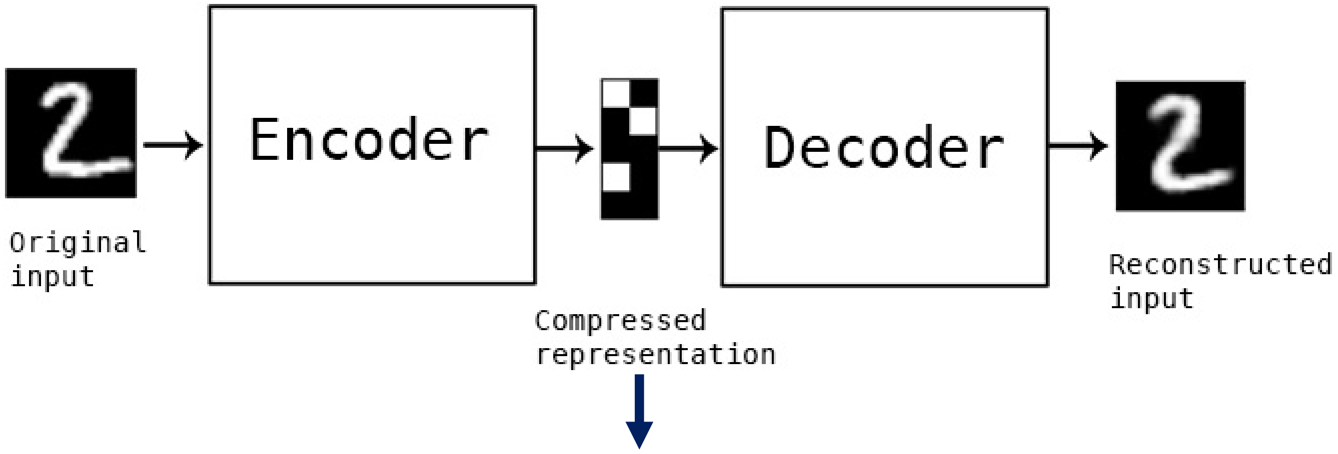
Deep learning in healthcare

Claim2vec: A new kind of feature

Medical codes are the essence of claims, but sparsity makes them difficult to use

- There are tens of thousands of codes that appear within claims
- Each claim line uses only a few codes

We train encoder and decoder neural networks to reconstruct codes used in claims



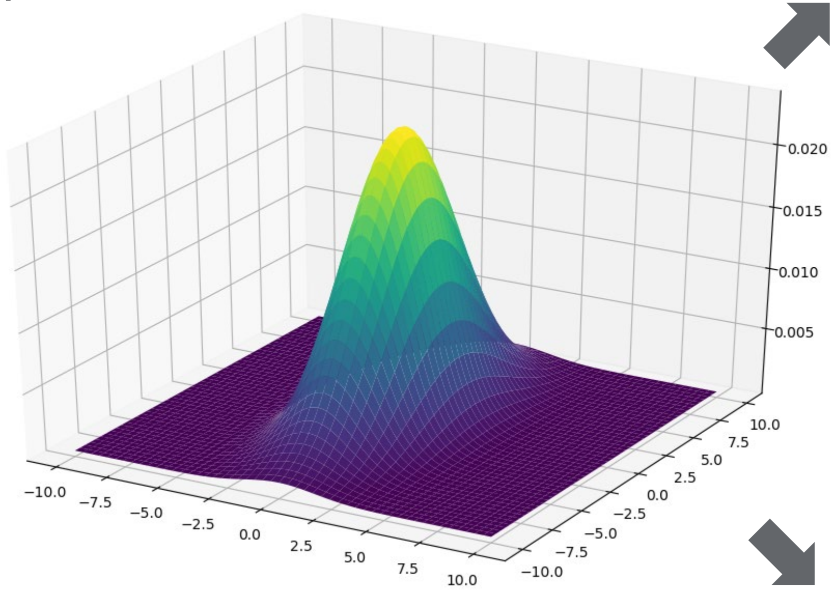
Extract claim2vec features

The compressed representation provides 100 dense claim2vec features for use in anomaly detection, and have increased our detection precision by 14%

fPSM: Facility predictive scoring model

Variational Autoencoder learned multivariate distribution for representative claims

DRG



High

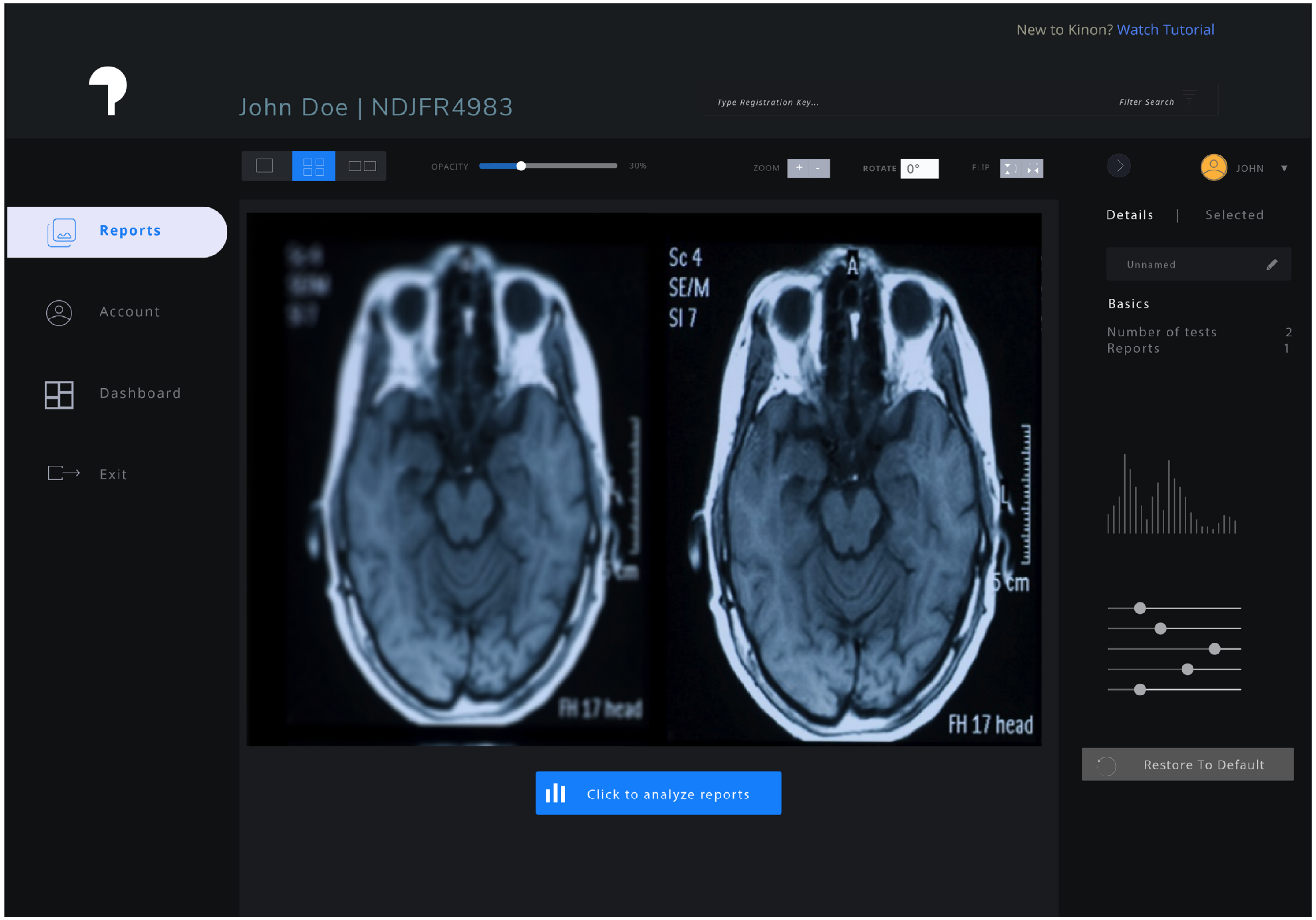
Reconstruction error

Low

Clinical input features, w/ facility claim embedding

GANs and medical imaging

- GANs that provide superior resolution to medical imaging.
- GANs are trained on several previously acquired high-resolution radiography images.
- The information from high-quality radiography imaging equipment can be leveraged even in areas with a relatively poor quality of the equipment.



https://medium.com/terenz/gans-in-healthcare-1a44f41c2d84

Summary

Why is deep learning so successful?

- Lots of data (examples)
  - 300 million+ photos are uploaded on Facebook EVERY DAY
  - 2.5 quintillion (18 zeros!) bytes of data generated every day
- Computing power
  - GPUs, TPUs, custom processors for specialty Deep Nets
  - able to build big and flexible models to translate and store examples
- Very flexible models
  - can handle numeric, nominal, categorical, image, video and sound data
  - able to build large and flexible models

From deep learning to deeply effective healthcare