Introduction to Machine Learning

Mark A. Austin

University of Maryland

austin@umd.edu

ENCE 688P, Fall Semester 2021

October 16, 2021

Overview

Part 03

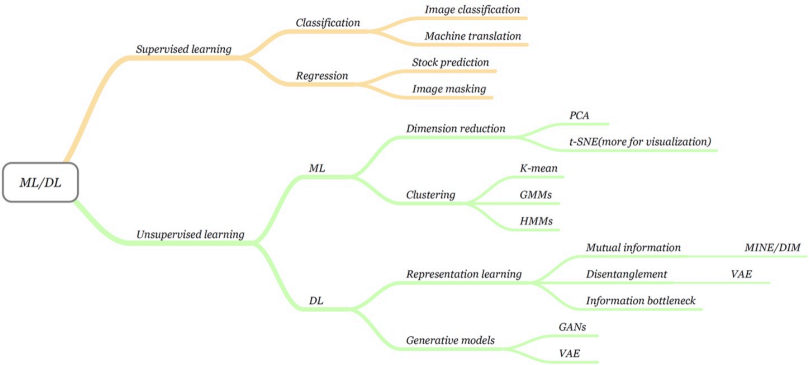
Taxonomy of Machine Learning Problems

Taxonomy of

Machine Learning Problems

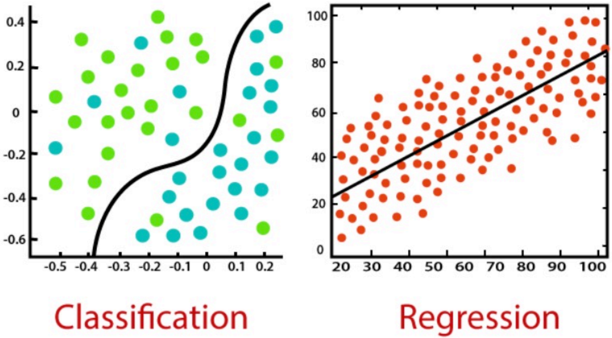
Classification of Machine Learning Problems

Tree of Machine Learning and Deep Learning Capabilities



Classification of Machine Learning Problems

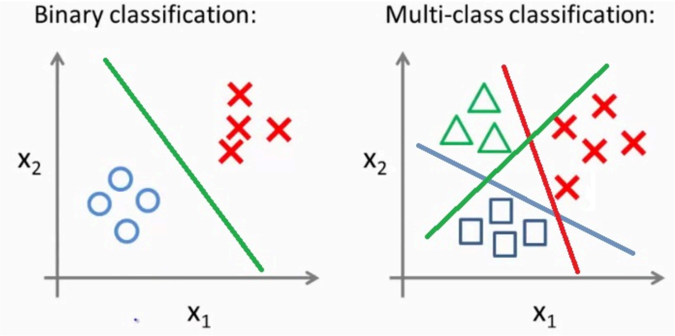
Regression versus Classification



Classification of Machine Learning Problems

Binary and Multi-Class Classification

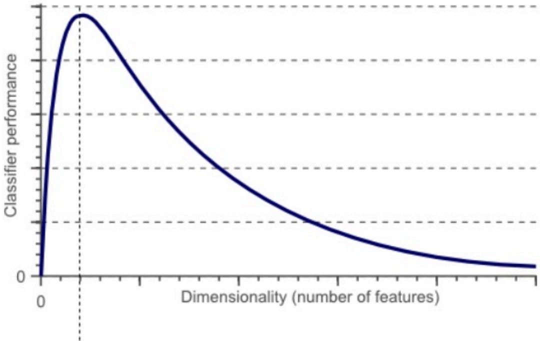
Task of separating elements of a set into two (or more) groups on the basis of a classification rule (e.g., shape, color, etc).



Classification of Machine Learning Problems

Curse of Dimensionality

Machine learning problems are inherently statistical and involve high-dimensional data. Increases in the problem dimensionality, decrease the number of data points available for classification in each dimension.

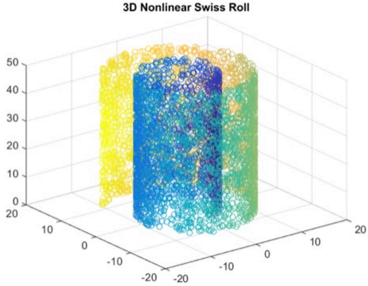


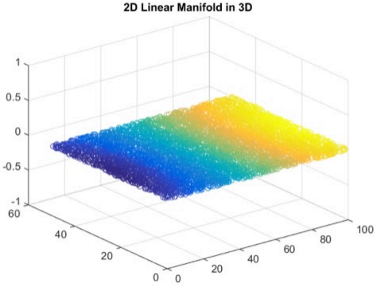
Classification of Machine Learning Problems

Dimensionality Reduction

Strategies of dimensionality reduction involve transformation of data to new (lower) dimension in such a way that some of the dimensions can be discarded without a loss of information.

Example: Projection of Swiss Roll data in 3D to 2D ...





Classification of Machine Learning Problems

Autoencoders

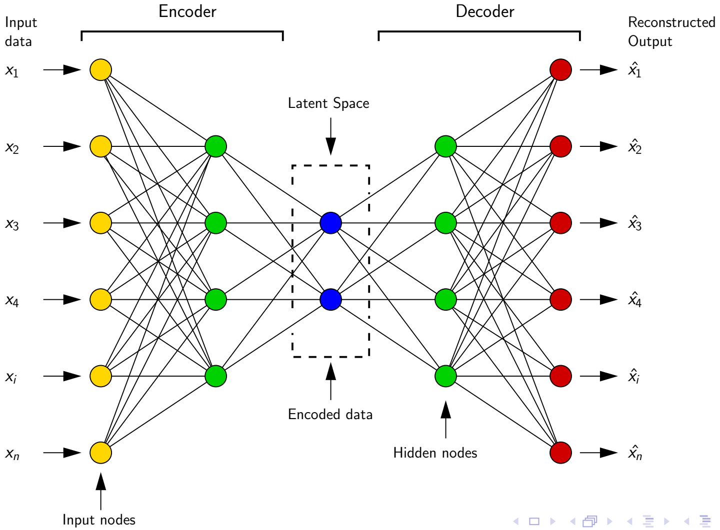
Autoencoder neural networks use unsupervised machine learning algorithms to: (1) find compressed representations of the input data (encoder), and (2) reconstruct the original data from the compressed data (decoder).

Applications:

- Dimensionality reduction.
- Image processing (compression and denoising).
- Feature extraction; anomaly detection.
- Image generation.
- Sequence-to-sequence translation.
- Recommendation systems.

Classification of Machine Learning Problems

AutoEncoder (Encoder-Decoder-Reconstruction)



Classification of Machine Learning Problems

Encoder

The encoder learns how to reduce the input dimensions and compress the input data into an encoded representation.

Decoder

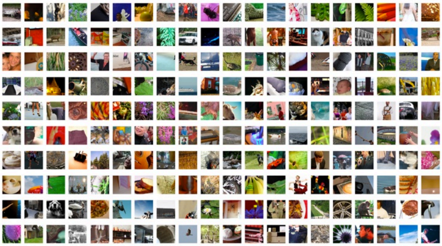
The decoder learns how to reconstruct the input data from the encoded representation and be as close to the input data as possible.

Latent Space

Latent space is simply a representation of compressed data in which similar points are closer together in space. This formalism is useful for learning data features and finding similar representations of data for analysis.

Classification of Machine Learning Problems

ImageNet and Deep Learning (2009-present)



Indexed Database of 14.2 million Images

- Project initiated by Fei Fei Li in 2006
- Image annotation process crowd sourced via Amazon's Mechanical Turk. Categories derived from WordNet.
- Well organized $\rightarrow$ supervised machine learning.

Classification of Machine Learning Problems

ImageNet and Deep Learning Capabilities:

- Identify objects in an image.
- 27 high-level categories; 21,800 sub-categories.

ImageNet and Deep Learning

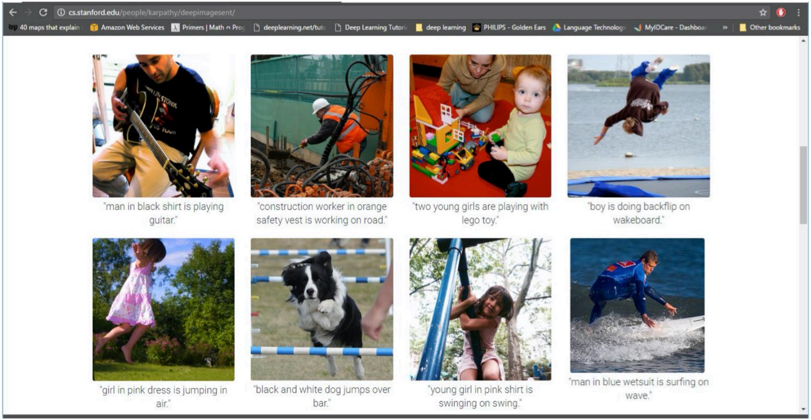
Capabilities (2018):

- Identify relationship among multiple objects in a image.

Example. Dog riding skateboard

ImageNet and Deep Learning

Captions generated by a neural network:



Machine Learning at Scale

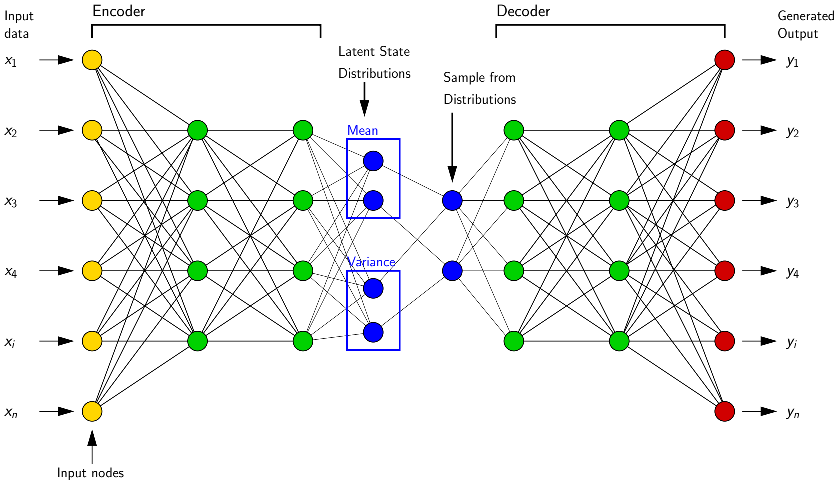
Object-recognition module:

- 24 million nodes.
- 140 million parameters.
- 15 billion connections.

Source: Fei Fei Li, TEDTalk, YouTube 2015.

Classification of Machine Learning Problems

Variational AutoEncoders (Generative Models)

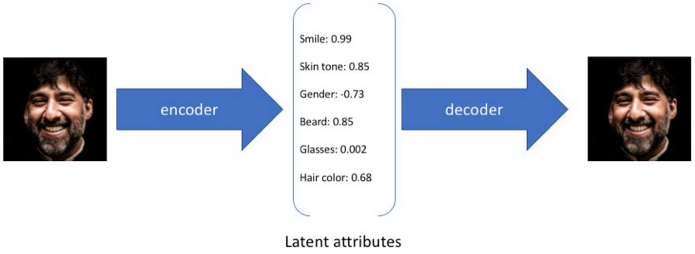


Classification of Machine Learning Problems

Standard Autoencoders vs. Variational Autoencoders:

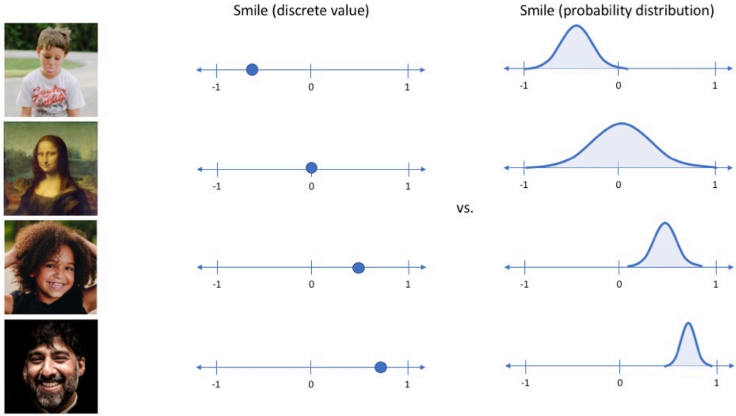
- A standard autoencoder outputs a single value for each encoding dimension.
- Variational autoencoders provide a probability distribution for each latent attribute.

Example: Single value representations for latent attributes:



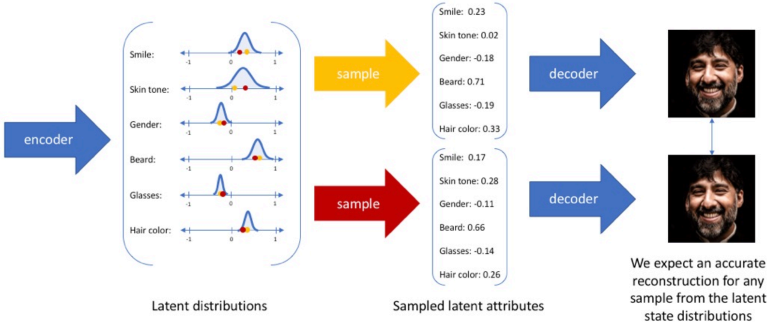
Classification of Machine Learning Problems

Discrete Value and Probability Distribution: Representations for smile latent attribute:



Classification of Machine Learning Problems

Image Reconstruction: sampled from latent distributions ...



Source: Jordan J., Variational Autoencoders, Data Science, March 2018.

References

- Austin M.A., Delgoshaei P., Coelho M. and Heidarinejad M. , Architecting Smart City Digital Twins: Combined Semantic Model and Machine Learning Approach, Journal of Management in Engineering, ASCE, Volume 36, Issue 4, July, 2020.
- Coelho M., and Austin M.A. , Teaching Machines to Understand Urban Networks, The Fifteenth International Conference on Systems (ICONS 2020), Lisbon, Portugal, February 23-27, 2020, pp. 37-42.
- Bhiksha R., Introduction to Neural Networks, Lisbon Machine Learning School, June, 2018.
- Lu T., Fundamental Limitations of Semi-Supervised Learning, MS Thesis in Mathematics in Computer Science, University of Waterloo, Canada, 2009.
- Van Engelen J.E., and Hoos H.H., A Survey on Semi-Supervised Learning, Machine Learning, Vol. 109, 2020, pp. 373-440.