

# March 13-17: Advanced machine learning and data analysis for the physical sciences

Morten Hjorth-Jensen<sup>1,2</sup>

<sup>1</sup>Department of Physics and Center for Computing in Science Education, University of Oslo, Norway

<sup>2</sup>Department of Physics and Astronomy and Facility for Rare Isotope Beams, Michigan State University, East Lansing, Michigan, USA

March 13-17, 2023

## Plans for the week March 13-17

- Summary of RNNs and discussion of Long-Short-Term memory with examples
- Discussion of Autoencoders (AEs)
- Reading recommendations:
  1. For RNNs see Goodfellow et al chapter 10. See also chapter 11 and 12 on practicalities and applications
  2. Reading suggestions for implementation of RNNs: [Aurelien Geron's chapter 14](#).
  3. [Deep Learning Tutorial on AEs from Stanford University](#)
  4. [Building AEs in Keras](#)
  5. [Introduction to AEs in TensorFlow](#)
  6. [Grosse, University of Toronto, Lecture on AEs](#)
  7. [Bank et al on AEs](#)

## Summary of RNNs

Recurrent neural networks (RNNs) have in general no probabilistic component in a model. With a given fixed input and target from data, the RNNs learn the intermediate association between various layers. The inputs, outputs, and internal representation (hidden states) are all real-valued vectors.

In a traditional NN, it is assumed that every input is independent of each other. But with sequential data, the input at a given stage  $t$  depends on the input from the previous stage  $t - 1$

### A typical RNN

1. Weight matrices  $U$ ,  $W$  and  $V$  that connect the input layer at a stage  $t$  with the hidden layer  $h_t$ , the previous hidden layer  $h_{t-1}$  with  $h_t$  and the hidden layer  $h_t$  connecting with the output layer at the same stage and producing an output  $\tilde{y}_t$ , respectively.
2. The output from the hidden layer  $h_t$  is often modulated by a tanh function  $h_t = f(x_t, h_{t-1}) = \tanh(Ux_t + Wh_{t-1} + b)$  with  $b$  a bias value
3. The output from the hidden layer produces  $\tilde{y}_t = g(Vh_t + c)$  where  $c$  is a new bias parameter.
4. The output from the training at a given stage is in turn compared with the observation  $y_t$  through a chosen cost function.

The function  $g$  can be any of the standard activation functions, that is a Sigmoid, a Softmax, a ReLU and other. The parameters are trained through the so-called back-propagation through time (BPTT) algorithm.

### Gating mechanism: Long Short Term Memory (LSTM)

Besides a simple recurrent neural network layer, as discussed above, there are two other commonly used types of recurrent neural network layers: Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU). For a short introduction to these layers see <https://medium.com/mindboard/lstm-vs-gru-experimental-comparison-955820c21e8b> and <https://medium.com/mindboard/lstm-vs-gru-experimental-comparison-955820c21e8b>.

LSTM uses a memory cell for modeling long-range dependencies and avoid vanishing gradient problems. Capable of modeling longer term dependencies by having memory cells and gates that controls the information flow along with the memory cells.

1. Introduced by Hochreiter and Schmidhuber (1997) who solved the problem of getting an RNN to remember things for a long time (like hundreds of time steps).
2. They designed a memory cell using logistic and linear units with multiplicative interactions.

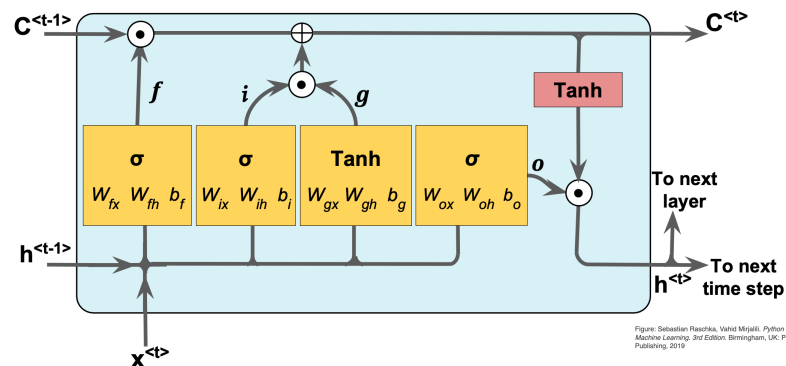
3. Information gets into the cell whenever its “write” gate is on.
4. The information stays in the cell so long as its **keep** gate is on.
5. Information can be read from the cell by turning on its **read** gate.

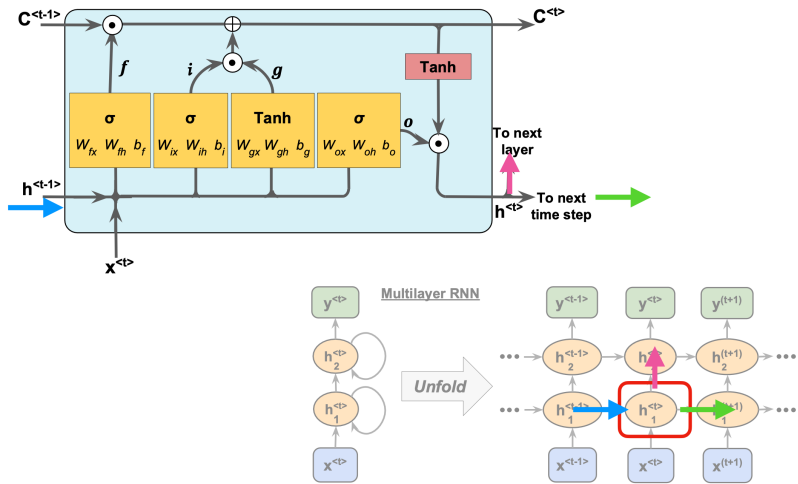
**Implementing a memory cell in a neural network.** To preserve information for a long time in the activities of an RNN, we use a circuit that implements an analog memory cell.

1. A linear unit that has a self-link with a weight of 1 will maintain its state.
2. Information is stored in the cell by activating its write gate.
3. Information is retrieved by activating the read gate.
4. We can backpropagate through this circuit because logistics are have nice derivatives.

## Long-short term memory (LSTM)

LSTM cell:

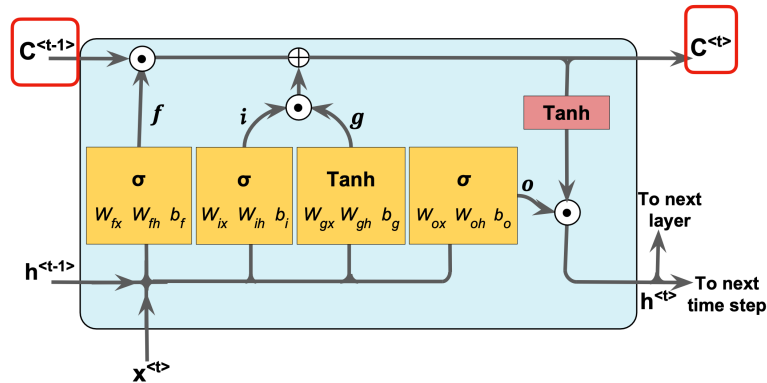




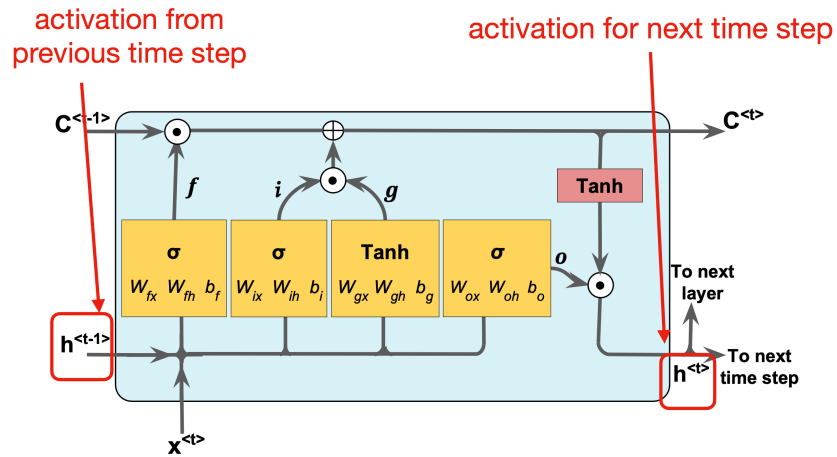
## Long-short term memory (LSTM)

Cell state at previous time step

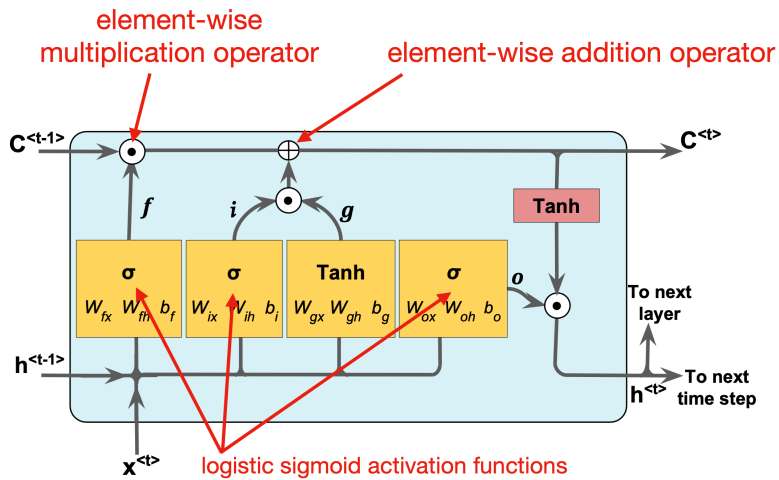
Cell state at current time step



## Long-short term memory (LSTM)



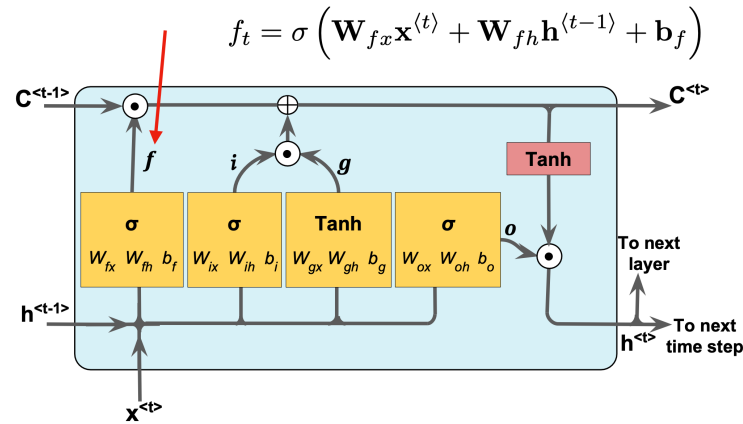
## Long-short term memory (LSTM)



## Long-short term memory (LSTM)

Gers, Felix A., Jürgen Schmidhuber, and Fred Cummins. "Learning to forget: Continual prediction with LSTM." (1999): 850-855.

**"Forget Gate":** controls which information is remembered, and which is forgotten; can reset the cell state

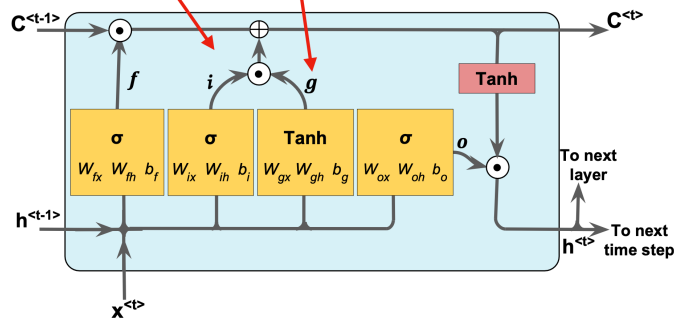


## Long-short term memory (LSTM)

**"Input Gate":**  $i_t = \sigma(W_{ix}x^{(t)} + W_{ih}h^{(t-1)} + b_i)$

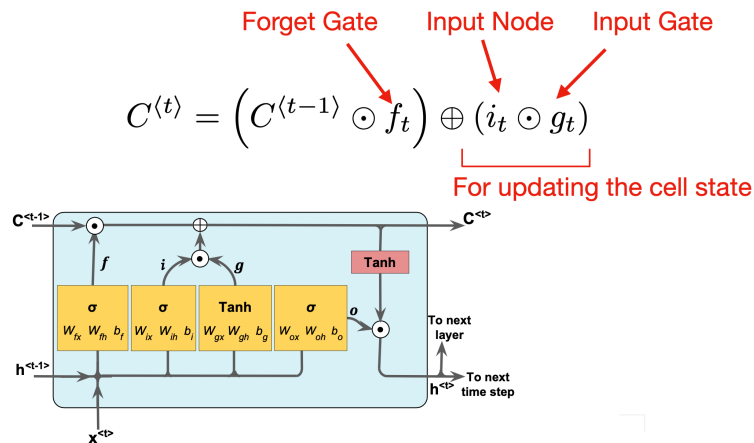
**"Input Node":**

$$g_t = \tanh(W_{gx}x^{(t)} + W_{gh}h^{(t-1)} + b_g)$$



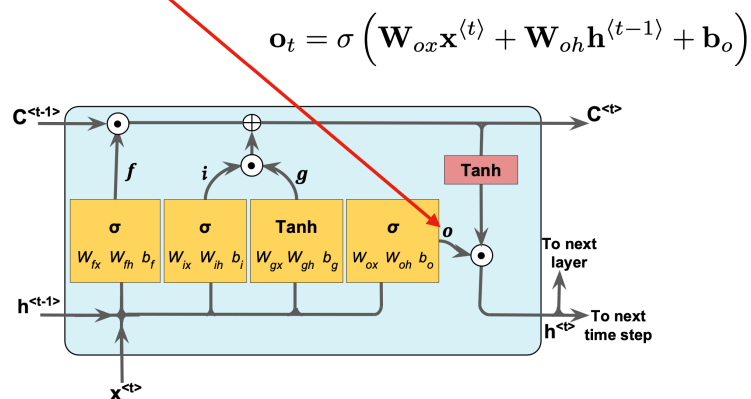
## Long-short term memory (LSTM)

Brief summary of the gates so far ...

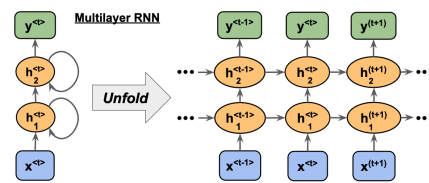
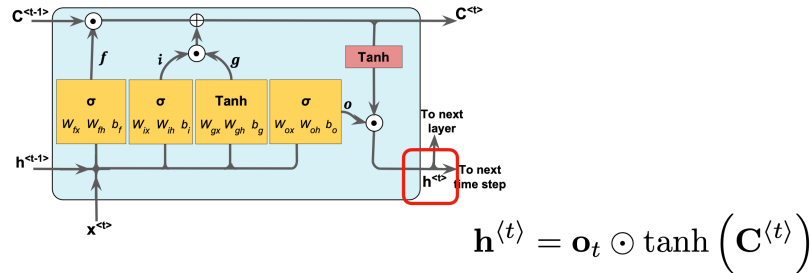


## Long-short term memory (LSTM)

Output gate for updating the values of hidden units:



## Long-short term memory (LSTM)



## An extrapolation example

The following code provides an example of how recurrent neural networks can be used to extrapolate to unknown values of physics data sets. Specifically, the data sets used in this program come from a quantum mechanical many-body calculation of energies as functions of the number of particles.

```
# For matrices and calculations
import numpy as np
# For machine learning (backend for keras)
import tensorflow as tf
# User-friendly machine learning library
# Front end for TensorFlow
import tensorflow.keras
# Different methods from Keras needed to create an RNN
# This is not necessary but it shortened function calls
# that need to be used in the code.
from tensorflow.keras import datasets, layers, models
from tensorflow.keras.layers import Input
from tensorflow.keras import regularizers
from tensorflow.keras.models import Model, Sequential
from tensorflow.keras.layers import Dense, SimpleRNN, LSTM, GRU
# For timing the code
from timeit import default_timer as timer
# For plotting
import matplotlib.pyplot as plt

# The data set
datatype='VaryDimension'
X_tot = np.arange(2, 42, 2)
```



```
y_tot = np.array([-0.03077640549, -0.08336233266, -0.1446729567, -0.2116753732, -0.2830637392, -0.3543561147, -0.4256484902, -0.4969408657, -0.5682332412, -0.6395256167, -0.7108179922, -0.7821103677, -0.8534027432, -0.9246951187, -0.9959874942, -1.0672798697, -1.1385722452, -1.2098646207, -1.2811569962, -1.3524493717, -1.4237417472, -1.4950341227, -1.5663264982, -1.6376188737, -1.7089112492, -1.7802036247, -1.8514959999, -1.9227883754, -1.9940807509, -2.0653731264, -2.1366655019, -2.2079578774, -2.2792502529, -2.3505426284, -2.4218350039, -2.4931273794, -2.5644197549, -2.6357121304, -2.7070045059, -2.7782968814, -2.8495892569, -2.9208816324, -2.9921740079, -3.0634663834, -3.1347587589, -3.2060511344, -3.2773435099, -3.3486358854, -3.4199282609, -3.4912206364, -3.5625130119, -3.6338053874, -3.7050977629, -3.7763901384, -3.8476825139, -3.9189748894, -3.9902672649, -4.0615596404, -4.1328520159, -4.2041443914, -4.2754367669, -4.3467291424, -4.4180215179, -4.4893138934, -4.5606062689, -4.6318986444, -4.7031910199, -4.7744833954, -4.8457757709, -4.9170681464, -4.9883605219, -5.0596528974, -5.1309452729, -5.2022376484, -5.2735300239, -5.3448223994, -5.4161147749, -5.4874071504, -5.5586995259, -5.6299919014, -5.7012842769, -5.7725766524, -5.8438690279, -5.9151614034, -5.9864537789, -6.0577461544, -6.1290385299, -6.2003309054, -6.2716232809, -6.3429156564, -6.4142080319, -6.4855004074, -6.5567927829, -6.6280851584, -6.6993775339, -6.7706699094, -6.8419622849, -6.9132546604, -6.9845470359, -7.0558394114, -7.1271317869, -7.1984241624, -7.2697165379, -7.3410089134, -7.4123012889, -7.4835936644, -7.5548860399, -7.6261784154, -7.6974707909, -7.7687631664, -7.8400555419, -7.9113479174, -7.9826402929, -8.0539326684, -8.1252250439, -8.1965174194, -8.2678097949, -8.3391021704, -8.4103945459, -8.4816869214, -8.5529792969, -8.6242716724, -8.6955640479, -8.7668564234, -8.8381487989, -8.9094411744, -8.9807335499, -9.0520259254, -9.1233183009, -9.1946106764, -9.2659030519, -9.3371954274, -9.4084878029, -9.4797801784, -9.5510725539, -9.6223649294, -9.6936573049, -9.7649496804, -9.8362420559, -9.9075344314, -9.9788268069, -10.0501191824, -10.1214115579, -10.1927039334, -10.2640063089, -10.3352986844, -10.4065910599, -10.4778834354, -10.5491758109, -10.6204681864, -10.6917605619, -10.7630529374, -10.8343453129, -10.9056376884, -10.9769300639, -11.0482224394, -11.1195148149, -11.1908071904, -11.2620995659, -11.3333919414, -11.4046843169, -11.4759766924, -11.5472690679, -11.6185614434, -11.6898538189, -11.7611461944, -11.8324385699, -11.9037309454, -11.9750233209, -12.0463156964, -12.1176080719, -12.1889004474, -12.2601928229, -12.3314851984, -12.4027775739, -12.4740699494, -12.5453623249, -12.6166547004, -12.6879470759, -12.7592394514, -12.8305318269, -12.9018242024, -12.9731165779, -13.0444089534, -13.1157013289, -13.1869937044, -13.2582860799, -13.3295784554, -13.4008708309, -13.4721632064, -13.5434555819, -13.6147479574, -13.6860403329, -13.7573327084, -13.8286250839, -13.8999174594, -13.9712098349, -14.0425022104, -14.1137945859, -14.1850869614, -14.2563793369, -14.3276717124, -14.3989640879, -14.4702564634, -14.5415488389, -14.6128412144, -14.6841335899, -14.7554259654, -14.8267183409, -14.8980107164, -14.9693030919, -15.0405954674, -15.1118878429, -15.1831802184, -15.2544725939, -15.3257649694, -15.3970573449, -15.4683497204, -15.5396420959, -15.6109344714, -15.6822268469, -15.7535192224, -15.8248115979, -15.8961039734, -15.9673963489, -16.0386887244, -16.1099810999, -16.1812734754, -16.2525658509, -16.3238582264, -16.3951506019, -16.4664429774, -16.5377353529, -16.6090277284, -16.6803201039, -16.7516124794, -16.8229048549, -16.8941972304, -16.9654896059, -17.0367819814, -17.1080743569, -17.1793667324, -17.2506591079, -17.3219514834, -17.3932438589, -17.4645362344, -17.5358286099, -17.6071209854, -17.6784133609, -17.7497057364, -17.8210081119, -17.8923004874, -17.9635928629, -18.0348852384, -18.1061776139, -18.1774699894, -18.2487623649, -18.3200547404, -18.3913471159, -18.4626394914, -18.5339318669, -18.6052242424, -18.6765166179, -18.7478089934, -18.8191013689, -18.8903937444, -18.9616861199, -19.0329784954, -19.1042708709, -19.1755632464, -19.2468556219, -19.3181479974, -19.3894403729, -19.4607327484, -19.5320251239, -19.6033174994, -19.6746098749, -19.7459022504, -19.8171946259, -19.8884870014, -19.9597793769, -20.0310717524, -20.1023641279, -20.1736565034, -20.2449488789, -20.3162412544, -20.3875336299, -20.4588260054, -20.5301183809, -20.6014107564, -20.6727031319, -20.7440055074, -20.8152978829, -20.8865902584, -20.9578826339, -21.0291750094, -21.1004673849, -21.1717597604, -21.2430521359, -21.3143445114, -21.3856368869, -21.4569292624, -21.5282216379, -21.5995140134, -21.6708063889, -21.7420987644, -21.8133911399, -21.8846835154, -21.9559758909, -22.0272682664, -22.0985606419, -22.1698530174, -22.2411453929, -22.3124377684, -22.3837301439, -22.4550225194, -22.5263148949, -22.5976072704, -22.6688996459, -22.7401920214, -22.8114843969, -22.8827767724, -22.9540691479, -23.0253615234, -23.0966538989, -23.1679462744, -23.2392386499, -23.3105310254, -23.3818234009, -23.4531157764, -23.5244081519, -23.5957005274, -23.6669929029, -23.7382852784, -23.8095776539, -23.8808700294, -23.9521624049, -24.0234547804, -24.0947471559, -24.1660395314, -24.2373319069, -24.3086242824, -24.3799166579, -24.4512090334, -24.5225014089, -24.5937937844, -24.6650861599, -24.7363785354, -24.8076709109, -24.8789632864, -24.9502556619, -25.0215480374, -25.0928404129, -25.1641327884, -25.2354251639, -25.3067175394, -25.3780099149, -25.4493022904, -25.5205946659, -25.5918870414, -25.6631794169, -25.7344717924, -25.8057641679, -25.8770565434, -25.9483489189, -26.0196412944, -26.0909336699, -26.1622260454, -26.2335184209, -26.3048107964, -26.3761031719, -26.4473955474, -26.5186879229, -26.5899802984, -26.6612726739, -26.7325650494, -26.8038574249, -26.8751498004, -26.9464421759, -27.0177345514, -27.0890269269, -27.1603193024, -27.2316116779, -27.3029040534, -27.3741964289, -27.4454888044, -27.5167811799, -27.5880735554, -27.6593659309, -27.7306583064, -27.8019506819, -27.8732430574, -27.9445354329, -28.0158278084, -28.0871201839, -28.1584125594, -28.2297049349, -28.3010073104, -28.3722996859, -28.4435920614, -28.5148844369, -28.5861768124, -28.6574691879, -28.7287615634, -28.8000539389, -28.8713463144, -28.9426386899, -29.0139310654, -29.0852234409, -29.1565158164, -29.2278081919, -29.2991005674, -29.3703929429, -29.4416853184, -29.5129776939, -29.5842700694, -29.6555624449, -29.7268548204, -29.7981471959, -29.8694395714, -29.9407319469, -30.0120243224, -30.0833166979, -30.1546090734, -30.2259014489, -30.2971938244, -30.3684861999, -30.4397785754, -30.5110709509, -30.5823633264, -30.6536557019, -30.7249480774, -30.7962404529, -30.8675328284, -30.9388252039, -31.0101175794, -31.0814099549, -31.1527023304, -31.2239947059, -31.2952870814, -31.3665794569, -31.4378718324, -31.5091642079, -31.5804565834, -31.6517489589, -31.7230413344, -31.7943337099, -31.8656260854, -31.9369184609, -32.0082108364, -32.0795032119, -32.1507955874, -32.2220879629, -32.2933803384, -32.3646727139, -32.4359650894, -32.5072574649, -32.5785498404, -32.6498422159, -32.7211345914, -32.7924269669, -32.8637193424, -32.9350117179, -33.0063040934, -33.0775964689, -33.1488888444, -33.2201812199, -33.2914735954, -33.3627659709, -33.4340583464, -33.5053507219, -33.5766430974, -33.6479354729, -33.7192278484, -33.7905202239, -33.8618125994, -33.9331049749, -34.0043973504, -34.0756897259, -34.1469821014, -34.2182744769, -34.2895668524, -34.3608592279, -34.4321516034, -34.5034439789, -34.5747363544, -34.6460287299, -34.7173211054, -34.7886134809, -34.8599058564, -34.9311982319, -35.0024906074, -35.0737829829, -35.1450753584, -35.2163677339, -35.2876601094, -35.3589524849, -35.4302448604, -35.5015372359, -35.5728296114, -35.6441219869, -35.7154143624, -35.7867067379, -35.8580091134, -35.9293014889, -36.0005938644, -36.0718862399, -36.1431786154, -36.2144709909, -36.2857633664, -36.3570557419, -36.4283481174, -36.4996404929, -36.5709328684, -36.6422252439, -36.7135176194, -36.7848100949, -36.8561024704, -36.9273948459, -37.0086872214, -37.0799795969, -37.1512719724, -37.2225643479, -37.2938567234, -37.3651490989, -37.4364414744, -37.5077338499, -37.5790262254, -37.6503186009, -37.7216109764, -37.7929033519, -37.8641957274, -37.9354881029, -38.0067804784, -38.0780728539, -38.1493652294, -38.2206576049, -38.2919499804, -38.3632423559, -38.4345347314, -38.5058271069, -38.5771194824, -38.6484118579, -38.7197042334, -38.7910066089, -38.8622989844, -38.9335913599, -39.0048837354, -39.0761761109, -39.1474684864, -39.2187608619, -39.2900532374, -39.3613456129, -39.4326379884, -39.5039303639, -39.5752227394, -39.6465151149, -39.7178074904, -39.7890998659, -39.8603922414, -39.9316846169, -40.0029769924, -40.0742693679, -40.1455617434, -40.2168541189, -40.2881464944, -40.3594388699, -40.4307312454, -40.5020236209, -40.5733159964, -40.6446083719, -40.7159007474, -40.7871931229, -40.8584854984, -40.9297778739, -41.0010702494, -41.0723626249, -41.1436550004, -41.2149473759, -41.2862397514, -41.3575321269, -41.4288245024, -41.5001168779, -41.5714092534, -41.6427016289, -41.7139940044, -41.7852863799, -41.8565787554, -41.9278711309, -42.0091635064, -42.0804558819, -42.1517482574, -42.2230406329, -42.2943330084, -42.3656253839, -42.4369177594, -42.5082101349, -42.5795025104, -42.6507948859, -42.7220872614, -42.7933796369, -42.8646720124, -42.9359643879, -43.0072567634, -43.0785491389, -43.1498415144, -43.2211338899, -43.2924262654, -43.3637186409, -43.4350110164, -43.5063033919, -43.5775957674, -43.6488881429, -43.7201805184, -43.7914728939, -43.8627652694, -43.9340576449, -44.0053500204, -44.0766423959, -44.1479347714, -44.2192271469, -44.2905195224, -44.3618118979, -44.4331042734, -44.5043966489, -44.5756890244, -44.6469813999, -44.7182737754, -44.7895661509, -44.8608585264, -44.9321509019, -45.0034432774, -45.0747356529, -45.1460280284, -45.2173204039, -45.2886127794, -45.3599051549, -45.4311975304, -45.5024899059, -45.5737822814, -45.6450746569, -45.7163670324, -45.7876594079, -45.8589517834, -45.9302441589, -46.0015365344, -46.0728289099, -46.1441212854, -46.2154136609, -46.2867060364, -46.3580084119, -46.4293007874, -46.5005931629, -46.5718855384, -46.6431779139, -46.7144702894, -46.7857626649, -46.8570550404, -46.9283474159, -47.0096397914, -47.0809321669, -47.1522245424, -47.2235169179, -47.2948092934, -47.3661016689, -47.4373940444, -47.5086864199, -47.5799787954, -47.6512711709, -47.7225635464, -47.7938559219, -47.8651482974, -47.9364406729, -48.0077330484, -48.0790254239, -48.1503177994, -48.2216101749, -48.2929025504, -48.3641949259, -48.4354873014, -48.5067796769, -48.5780720524, -48.6493644279, -48.7206568034, -48.7919491789, -48.8632415544, -48.9345339299, -49.0058263054, -49.0771186809, -49.1484110564, -49.2197034319, -49.2910058074, -49.3622981829, -49.4335905584, -49.5048829339, -49.5761753094, -49.6474676849, -49.7187600604, -49.7900524359, -49.8613448114, -49.9326371869, -50.0039295624, -50.0752219379, -50.1465143134, -50.2178066889, -50.2890990644, -50.3603914399, -50.4316838154, -50.5029761909, -50.5742685664, -50.6455609419, -50.7168533174, -50.7881456929, -50.8594380684, -50.9307304439, -51.0020228194, -51.0733151949, -51.1446075704, -51.2158999459, -51.2871923214, -51.3584846969, -51.4297770724, -51.5010694479, -51.5723618234, -51.6436541989, -51.7149465744, -51.7862389499, -51.8575313254, -51.9288237009, -52.0001160764, -52.0714084519, -52.1427008274, -52.2139932029, -52.2852855784, -52.3565779539, -52.4278703294, -52.4991627049, -52.5704550804, -52.6417474559, -52.7130398314, -52.7843322069, -52.8556245824, -52.9269169579, -53.0082093334, -53.0795017089, -53.1507940844, -53.2220864599, -53.2933788354, -53.3646712109, -53.4359635864, -53.5072559619, -53.5785483374, -53.6498407129, -53.7211330884, -53.7924254639, -53.8637178394, -53.9350102149, -54.0063025904, -54.0775949659, -54.1488873414, -54.2201797169, -54.2914720924, -54.3627644679, -54.4340568434, -54.5053492189, -54.5766415944, -54.6479339699, -54.7192263454, -54.7905187209, -54.8618110964, -54.9331034719, -55.0043958474, -55.0756882229, -55.1469805984, -55.2182729739, -55.2895653494, -55.3608577249, -55.4321501004, -55.5034424759, -55.5747348514, -55.6460272269, -55.7173196024, -55.7886119779, -55.8599043534, -55.9311967289, -56.0024891044, -56.0737814799, -56.1450738554, -56.2163662309, -56.2876586064, -56.3589509819, -56.4302433574, -56.5015357329, -56.5728281084, -56.6441204839, -56.7154128594, -56.7867052349, -56.8579976104, -56.9292899859, -57.0005823614, -57.0718747369, -57.1431671124, -57.2144594879, -57.2857518634, -57.3570442389, -57.4283366144, -57.4996289899, -57.5709213654, -57.6422137409, -57.7135061164, -57.7847984919, -57.8560908674, -57.9273832429, -58.0086756184, -58.0799679939, -58.1512603694, -58.2225527449, -58.2938451204, -58.3651374959, -58.4364298714, -58.5077222469, -58.5790146224, -58.6503069979, -58.7215993734, -58.7928917489, -58.8641841244, -58.9354765009, -59.0067688764, -59.0780612519, -59.1493536274, -59.2206460029, -59.2919383784, -59.363230
```

```

        a = np.reshape (a, (len(a), 1))
        X.append(a)
        Y.append(b)
    rnn_input = np.array(X)
    rnn_output = np.array(Y)

    return rnn_input, rnn_output

# ## Defining the Recurrent Neural Network Using Keras
#
# The following method defines a simple recurrent neural network in keras consisting of one input
def rnn(length_of_sequences, batch_size = None, stateful = False):
    """
    Inputs:
        length_of_sequences (an int): the number of y values in "x data". This is determined
            when the data is formatted
        batch_size (an int): Default value is None. See Keras documentation of SimpleRNN.
        stateful (a boolean): Default value is False. See Keras documentation of SimpleRNN.
    Returns:
        model (a Keras model): The recurrent neural network that is built and compiled by this
            method
    """
    Builds and compiles a recurrent neural network with one hidden layer and returns the model
    """
    # Number of neurons in the input and output layers
    in_out_neurons = 1
    # Number of neurons in the hidden layer
    hidden_neurons = 200
    # Define the input layer
    inp = Input(batch_shape=(batch_size,
                              length_of_sequences,
                              in_out_neurons))
    # Define the hidden layer as a simple RNN layer with a set number of neurons and add it to
    # the network immediately after the input layer
    rnn = SimpleRNN(hidden_neurons,
                    return_sequences=False,
                    stateful = stateful,
                    name="RNN")(inp)
    # Define the output layer as a dense neural network layer (standard neural network layer)
    # and add it to the network immediately after the hidden layer.
    dens = Dense(in_out_neurons,name="dense")(rnn)
    # Create the machine learning model starting with the input layer and ending with the
    # output layer
    model = Model(inputs=[inp],outputs=[dens])
    # Compile the machine learning model using the mean squared error function as the loss
    # function and an Adams optimizer.
    model.compile(loss="mean_squared_error", optimizer="adam")
    return model

```

## Predicting New Points With A Trained Recurrent Neural Network

```

def test_rnn (x1, y_test, plot_min, plot_max):
    """
    Inputs:
        x1 (a list or numpy array): The complete x component of the data set

```

```

        y_test (a list or numpy array): The complete y component of the data set
        plot_min (an int or float): the smallest x value used in the training data
        plot_max (an int or float): the largest x value used in the training data
    Returns:
        None.
    Uses a trained recurrent neural network model to predict future points in the
    series. Computes the MSE of the predicted data set from the true data set, saves
    the predicted data set to a csv file, and plots the predicted and true data sets w
    hile also displaying the data range used for training.
    """
    # Add the training data as the first dim points in the predicted data array as these
    # are known values.
    y_pred = y_test[:dim].tolist()
    # Generate the first input to the trained recurrent neural network using the last two
    # points of the training data. Based on how the network was trained this means that it
    # will predict the first point in the data set after the training data. All of the
    # brackets are necessary for Tensorflow.
    next_input = np.array([[y_test[dim-2]], [y_test[dim-1]]])
    # Save the very last point in the training data set. This will be used later.
    last = [y_test[dim-1]]

    # Iterate until the complete data set is created.
    for i in range(dim, len(y_test)):
        # Predict the next point in the data set using the previous two points.
        next = model.predict(next_input)
        # Append just the number of the predicted data set
        y_pred.append(next[0][0])
        # Create the input that will be used to predict the next data point in the data set.
        next_input = np.array([[last, next[0]]], dtype=np.float64)
        last = next

    # Print the mean squared error between the known data set and the predicted data set.
    print('MSE: ', np.square(np.subtract(y_test, y_pred)).mean())
    # Save the predicted data set as a csv file for later use
    name = datatype + 'Predicted'+str(dim)+'.csv'
    np.savetxt(name, y_pred, delimiter=',')
    # Plot the known data set and the predicted data set. The red box represents the region that
    # for the training data.
    fig, ax = plt.subplots()
    ax.plot(x1, y_test, label="true", linewidth=3)
    ax.plot(x1, y_pred, 'g-.', label="predicted", linewidth=4)
    ax.legend()
    # Created a red region to represent the points used in the training data.
    ax.axvspan(plot_min, plot_max, alpha=0.25, color='red')
    plt.show()

    # Check to make sure the data set is complete
    assert len(X_tot) == len(y_tot)

    # This is the number of points that will be used in as the training data
    dim=12

    # Separate the training data from the whole data set
    X_train = X_tot[:dim]
    y_train = y_tot[:dim]

    # Generate the training data for the RNN, using a sequence of 2
    rnn_input, rnn_training = format_data(y_train, 2)

```

```

# Create a recurrent neural network in Keras and produce a summary of the
# machine learning model
model = rnn(length_of_sequences = rnn_input.shape[1])
model.summary()

# Start the timer. Want to time training+testing
start = timer()
# Fit the model using the training data generated above using 150 training iterations and a 5%
# validation split. Setting verbose to True prints information about each training iteration.
hist = model.fit(rnn_input, rnn_training, batch_size=None, epochs=150,
                verbose=True, validation_split=0.05)

for label in ["loss", "val_loss"]:
    plt.plot(hist.history[label], label=label)

plt.ylabel("loss")
plt.xlabel("epoch")
plt.title("The final validation loss: {}".format(hist.history["val_loss"][-1]))
plt.legend()
plt.show()

# Use the trained neural network to predict more points of the data set
test_rnn(X_tot, y_tot, X_tot[0], X_tot[dim-1])
# Stop the timer and calculate the total time needed.
end = timer()
print('Time: ', end-start)

```

## Other Things to Try

Changing the size of the recurrent neural network and its parameters can drastically change the results you get from the model. The below code takes the simple recurrent neural network from above and adds a second hidden layer, changes the number of neurons in the hidden layer, and explicitly declares the activation function of the hidden layers to be a sigmoid function. The loss function and optimizer can also be changed but are kept the same as the above network. These parameters can be tuned to provide the optimal result from the network. For some ideas on how to improve the performance of a recurrent neural network.

```

def rnn_2layers(length_of_sequences, batch_size = None, stateful = False):
    """
    Inputs:
        length_of_sequences (an int): the number of y values in "x data". This is determined
            when the data is formatted
        batch_size (an int): Default value is None. See Keras documentation of SimpleRNN.
        stateful (a boolean): Default value is False. See Keras documentation of SimpleRNN.
    Returns:
        model (a Keras model): The recurrent neural network that is built and compiled by this
            method
    Builds and compiles a recurrent neural network with two hidden layers and returns the model
    """
    # Number of neurons in the input and output layers
    in_out_neurons = 1
    # Number of neurons in the hidden layer, increased from the first network
    hidden_neurons = 500
    # Define the input layer

```

```

inp = Input(batch_shape=(batch_size,
                        length_of_sequences,
                        in_out_neurons))
# Create two hidden layers instead of one hidden layer. Explicitly set the activation
# function to be the sigmoid function (the default value is hyperbolic tangent)
rnn1 = SimpleRNN(hidden_neurons,
                return_sequences=True, # This needs to be True if another hidden layer is to
                stateful = stateful, activation = 'sigmoid',
                name="RNN1")(inp)
rnn2 = SimpleRNN(hidden_neurons,
                return_sequences=False, activation = 'sigmoid',
                stateful = stateful,
                name="RNN2")(rnn1)
# Define the output layer as a dense neural network layer (standard neural network layer)
# and add it to the network immediately after the hidden layer.
dens = Dense(in_out_neurons,name="dense")(rnn2)
# Create the machine learning model starting with the input layer and ending with the
# output layer
model = Model(inputs=[inp],outputs=[dens])
# Compile the machine learning model using the mean squared error function as the loss
# function and an Adams optimizer.
model.compile(loss="mean_squared_error", optimizer="adam")
return model

# Check to make sure the data set is complete
assert len(X_tot) == len(y_tot)

# This is the number of points that will be used in as the training data
dim=12

# Separate the training data from the whole data set
X_train = X_tot[:dim]
y_train = y_tot[:dim]

# Generate the training data for the RNN, using a sequence of 2
rnn_input, rnn_training = format_data(y_train, 2)

# Create a recurrent neural network in Keras and produce a summary of the
# machine learning model
model = rnn_2layers(length_of_sequences = 2)
model.summary()

# Start the timer. Want to time training+testing
start = timer()
# Fit the model using the training data generated above using 150 training iterations and a 5%
# validation split. Setting verbose to True prints information about each training iteration.
hist = model.fit(rnn_input, rnn_training, batch_size=None, epochs=150,
                verbose=True,validation_split=0.05)

# This section plots the training loss and the validation loss as a function of training iteration
# This is not required for analyzing the couple cluster data but can help determine if the network
# being overtrained.
for label in ["loss","val_loss"]:
    plt.plot(hist.history[label],label=label)

plt.ylabel("loss")
plt.xlabel("epoch")
plt.title("The final validation loss: {}".format(hist.history["val_loss"][-1]))

```

```

plt.legend()
plt.show()

# Use the trained neural network to predict more points of the data set
test_rnn(X_tot, y_tot, X_tot[0], X_tot[dim-1])
# Stop the timer and calculate the total time needed.
end = timer()
print('Time: ', end-start)

```

## Other Types of Recurrent Neural Networks

The first network created below is similar to the previous network, but it replaces the SimpleRNN layers with LSTM layers. The second network below has two hidden layers made up of GRUs, which are preceded by two dense (feedforward) neural network layers. This architecture has been shown to improve the performance of recurrent neural networks (see the link above and also <https://arxiv.org/pdf/1807.02857.pdf>).

```

def lstm_2layers(length_of_sequences, batch_size = None, stateful = False):
    """
    Inputs:
        length_of_sequences (an int): the number of y values in "x data". This is determined
            when the data is formatted
        batch_size (an int): Default value is None. See Keras documentation of SimpleRNN.
        stateful (a boolean): Default value is False. See Keras documentation of SimpleRNN.
    Returns:
        model (a Keras model): The recurrent neural network that is built and compiled by this
            method
    Builds and compiles a recurrent neural network with two LSTM hidden layers and returns the
    """
    # Number of neurons on the input/output layer and the number of neurons in the hidden layer
    in_out_neurons = 1
    hidden_neurons = 250
    # Input Layer
    inp = Input(batch_shape=(batch_size,
                              length_of_sequences,
                              in_out_neurons))
    # Hidden layers (in this case they are LSTM layers instead of SimpleRNN layers)
    rnn = LSTM(hidden_neurons,
                return_sequences=True,
                stateful = stateful,
                name="RNN", use_bias=True, activation='tanh')(inp)
    rnn1 = LSTM(hidden_neurons,
                 return_sequences=False,
                 stateful = stateful,
                 name="RNN1", use_bias=True, activation='tanh')(rnn)
    # Output layer
    dens = Dense(in_out_neurons, name="dense")(rnn1)
    # Define the model
    model = Model(inputs=[inp], outputs=[dens])
    # Compile the model
    model.compile(loss='mean_squared_error', optimizer='adam')
    # Return the model
    return model

def dnn2_gru2(length_of_sequences, batch_size = None, stateful = False):

```

```

"""
    Inputs:
        length_of_sequences (an int): the number of y values in "x data". This is determined
            when the data is formatted
        batch_size (an int): Default value is None. See Keras documentation of SimpleRNN.
        stateful (a boolean): Default value is False. See Keras documentation of SimpleRNN.
    Returns:
        model (a Keras model): The recurrent neural network that is built and compiled by this
            method
    Builds and compiles a recurrent neural network with four hidden layers (two dense followed
        two GRU layers) and returns the model.
"""
# Number of neurons on the input/output layers and hidden layers
in_out_neurons = 1
hidden_neurons = 250
# Input layer
inp = Input(batch_shape=(batch_size,
                          length_of_sequences,
                          in_out_neurons))
# Hidden Dense (feedforward) layers
dnn = Dense(hidden_neurons/2, activation='relu', name='dnn')(inp)
dnn1 = Dense(hidden_neurons/2, activation='relu', name='dnn1')(dnn)
# Hidden GRU layers
rnn1 = GRU(hidden_neurons,
            return_sequences=True,
            stateful = stateful,
            name="RNN1", use_bias=True)(dnn1)
rnn = GRU(hidden_neurons,
            return_sequences=False,
            stateful = stateful,
            name="RNN", use_bias=True)(rnn1)
# Output layer
dens = Dense(in_out_neurons, name="dense")(rnn)
# Define the model
model = Model(inputs=[inp], outputs=[dens])
# Compile the model
model.compile(loss='mean_squared_error', optimizer='adam')
# Return the model
return model

# Check to make sure the data set is complete
assert len(X_tot) == len(y_tot)

# This is the number of points that will be used in as the training data
dim=12

# Separate the training data from the whole data set
X_train = X_tot[:dim]
y_train = y_tot[:dim]

# Generate the training data for the RNN, using a sequence of 2
rnn_input, rnn_training = format_data(y_train, 2)

# Create a recurrent neural network in Keras and produce a summary of the
# machine learning model
# Change the method name to reflect which network you want to use
model = dnn2_gru2(length_of_sequences = 2)
model.summary()

```

```

# Start the timer. Want to time training+testing
start = timer()
# Fit the model using the training data generated above using 150 training iterations and a 5%
# validation split. Setting verbose to True prints information about each training iteration.
hist = model.fit(rnn_input, rnn_training, batch_size=None, epochs=150,
                 verbose=True, validation_split=0.05)

# This section plots the training loss and the validation loss as a function of training iteration
# This is not required for analyzing the couple cluster data but can help determine if the network
# being overtrained.
for label in ["loss", "val_loss"]:
    plt.plot(hist.history[label], label=label)

plt.ylabel("loss")
plt.xlabel("epoch")
plt.title("The final validation loss: {}".format(hist.history["val_loss"][-1]))
plt.legend()
plt.show()

# Use the trained neural network to predict more points of the data set
test_rnn(X_tot, y_tot, X_tot[0], X_tot[dim-1])
# Stop the timer and calculate the total time needed.
end = timer()
print('Time: ', end-start)

# ### Training Recurrent Neural Networks in the Standard Way (i.e. learning the relationship between
#
# Finally, comparing the performance of a recurrent neural network using the standard data formatting

# Check to make sure the data set is complete
assert len(X_tot) == len(y_tot)

# This is the number of points that will be used in as the training data
dim=12

# Separate the training data from the whole data set
X_train = X_tot[:dim]
y_train = y_tot[:dim]

# Reshape the data for Keras specifications
X_train = X_train.reshape((dim, 1))
y_train = y_train.reshape((dim, 1))

# Create a recurrent neural network in Keras and produce a summary of the
# machine learning model
# Set the sequence length to 1 for regular data formatting
model = rnn(length_of_sequences = 1)
model.summary()

# Start the timer. Want to time training+testing
start = timer()
# Fit the model using the training data generated above using 150 training iterations and a 5%
# validation split. Setting verbose to True prints information about each training iteration.
hist = model.fit(X_train, y_train, batch_size=None, epochs=150,
                 verbose=True, validation_split=0.05)

# This section plots the training loss and the validation loss as a function of training iteration

```



```

# This is not required for analyzing the couple cluster data but can help determine if the network
# being overtrained.
for label in ["loss", "val_loss"]:
    plt.plot(hist.history[label], label=label)

plt.ylabel("loss")
plt.xlabel("epoch")
plt.title("The final validation loss: {}".format(hist.history["val_loss"][-1]))
plt.legend()
plt.show()

# Use the trained neural network to predict the remaining data points
X_pred = X_tot[dim:]
X_pred = X_pred.reshape((len(X_pred), 1))
y_model = model.predict(X_pred)
y_pred = np.concatenate((y_tot[:dim], y_model.flatten()))

# Plot the known data set and the predicted data set. The red box represents the region that was
# for the training data.
fig, ax = plt.subplots()
ax.plot(X_tot, y_tot, label="true", linewidth=3)
ax.plot(X_tot, y_pred, 'g-.', label="predicted", linewidth=4)
ax.legend()
# Created a red region to represent the points used in the training data.
ax.axvspan(X_tot[0], X_tot[dim], alpha=0.25, color='red')
plt.show()

# Stop the timer and calculate the total time needed.
end = timer()
print('Time: ', end-start)

```

## Autoencoders: Overarching view

Autoencoders are artificial neural networks capable of learning efficient representations of the input data (these representations are called codings) without any supervision (i.e., the training set is unlabeled). These codings typically have a much lower dimensionality than the input data, making autoencoders useful for dimensionality reduction.

More importantly, autoencoders act as powerful feature detectors, and they can be used for unsupervised pretraining of deep neural networks.

Lastly, they are capable of randomly generating new data that looks very similar to the training data; this is called a generative model. For example, you could train an autoencoder on pictures of faces, and it would then be able to generate new faces. Surprisingly, autoencoders work by simply learning to copy their inputs to their outputs. This may sound like a trivial task, but we will see that constraining the network in various ways can make it rather difficult. For example, you can limit the size of the internal representation, or you can add noise to the inputs and train the network to recover the original inputs. These constraints prevent the autoencoder from trivially copying the inputs directly to the outputs, which forces it to learn efficient ways of representing the data.

In short, the codings are byproducts of the autoencoder's attempt to learn the identity function under some constraints.

## From supervised to unsupervised training

Neural networks are typically used in a supervised setting. Meaning that for each training observation we will have one label or target/output value. During training, the neural network model will learn the relationship between the input data and the expected labels. Now suppose we have only unlabeled observations, meaning we only have our training dataset  $S_T$  made of the  $M$  observations  $\mathbf{x}_i$  with  $i = 1, \dots, M$

$$S_T = \{\mathbf{x}_i | i = 1, \dots, M\} \tag{1}$$

where in general  $\mathbf{x}_i \in \mathbb{R}^n$  with  $n \in \mathbb{N}$ .

## First introduction of AEs

Autoencoders were first introduced by Rumelhart, Hinton, and Williams in 1986 with the goal of learning to reconstruct the input observations  $\mathbf{x}_i$  with the lowest error possible.

Why would one want to learn to reconstruct the input observations? If you have problems imagining what that means, think of having a dataset made of images. An autoencoder would be an algorithm that can give as output an image that is as similar as possible to the input one. You may be confused, as there is no apparent reason of doing so. To better understand why autoencoders are useful we need a more informative (although not yet unambiguous) definition.

An autoencoder is a type of algorithm with the primary purpose of learning an "informative" representation of the data that can be used for different applications (see [Bank, D., Koenigstein, N., and Giryas, R., Autoencoders](#)) by learning to reconstruct a set of input observations well enough.

To better understand autoencoders we need to refer to their typical architecture, visualized in Figure (to be added). The autoencoders' main components are three: an encoder, a latent feature representation, and a decoder. The encoder and decoder are simply functions, while with the name **latent feature representation**, one typically intends a tensor of real numbers.