Rachael Girardey, Lauren Clarke, Alex Jones
Week 4 Course Notes

Deep Learning Architecture

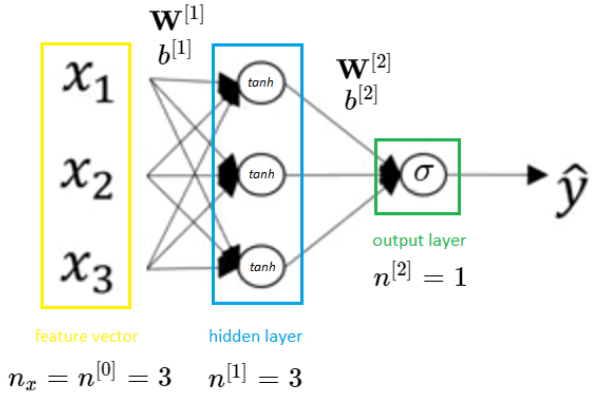
What is the primary difference between Deep Learning and Standard Machine Learning?

1. DL is a Representation Learning Method = people have traditionally collected data from various sources but these features aren't powerful to predict information from this data
   - To things we want the machine to learn, need features
   - Existing methods you need domain expertise to tell you what data and features are actually important
   - A ML model will look at the features to tell you the best possible outcome
     a. Example is determining the price of a house based on specific features
2. Standard ML does not have enough power to look at the multiple features
3. Standard ML works well with the raw data →
   a. To do this it: builds a feature extractor from domain experts (humans)
   b. Get complicated unintelligible features
   c. Feed into a standard ML model (such as SVM)
   d. Get outcome
   - Drawback: Feature extractor needs to be built specifically for the topic
4. DL is learning alternate and more useful visual representations of data
   a. Many layers instead of feature extractor (form useful representations for the problem) = automated the task = Representation Learning Methods
   b. Anybody can be a data scientist as long as they have good data

Process in the past:
SML- used to have raw data -> **feature extractor** (domain experts- hand crafted features- features from domain insights)->complicated, unintelligent features-> standard ML models SVM -> Output

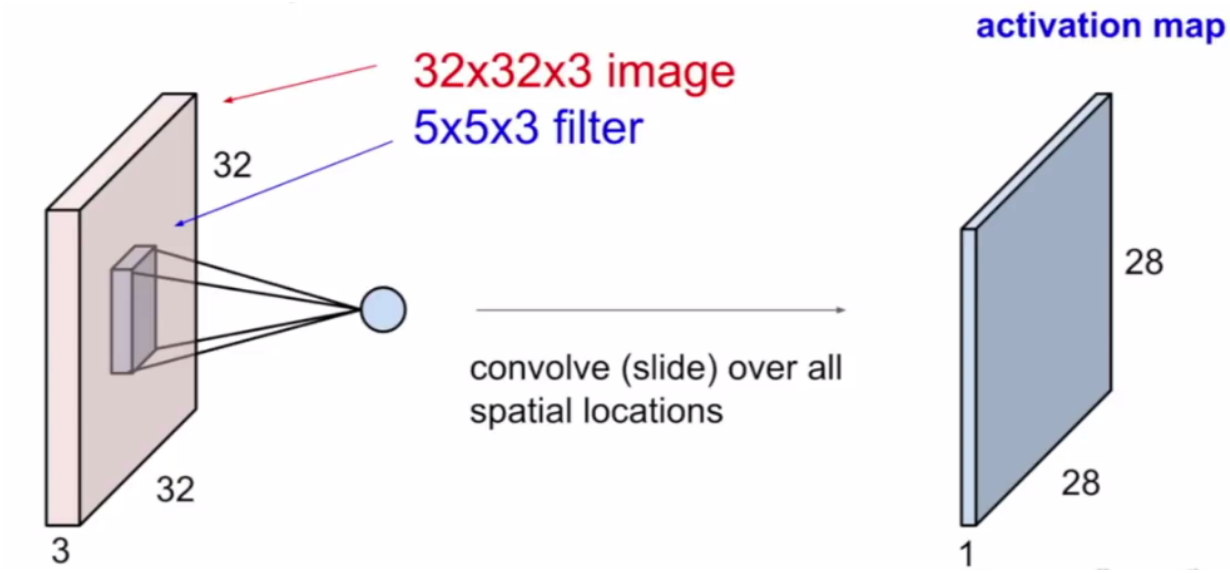
4 categories of DL models → use of them depends upon the problem, but there is no rule on which one you use

1. *Standard Vanilla Model*
   - Uses no particular structure just numerical forms → gradient descent
   - Each neuron is deciding its output (for the image)
   - Tabular form



Source: http://media5.datahacker.rs/2018/09/010_NN_v2.png

2. *Convolutional Neural Network (CNN)*
   - **Spatial dependence**
   - Exploiting the fact that as images form so is some sort of object (any image data, heat maps, CT scans)
   - In images, you have spatial context that are important to maintain → at the end of the computation you have a confidence scale of what the image is
   - Images have W, H, D (32 X 32 X 3) which are spatial structures
     - Each pixel will become a feature into a 3,072 dataset → if doing this, spatial information will be lost in translation (location)
     - Using the standard model and inputting specific information
   - Convolution layer
     - Convolve the filter with the image (slide over the image spatial computing dot products) = place the filter over the original image and giving you the next level features to see if it corresponds to a feature of the image (nose, eyes, neck, etc)
     - Filters = specific dimension of the image (3 dimensional)
     - 1 number: the result of taking a dot product between the filter and a small 5x5x3 chunk of the image (75-dimensional dot product)

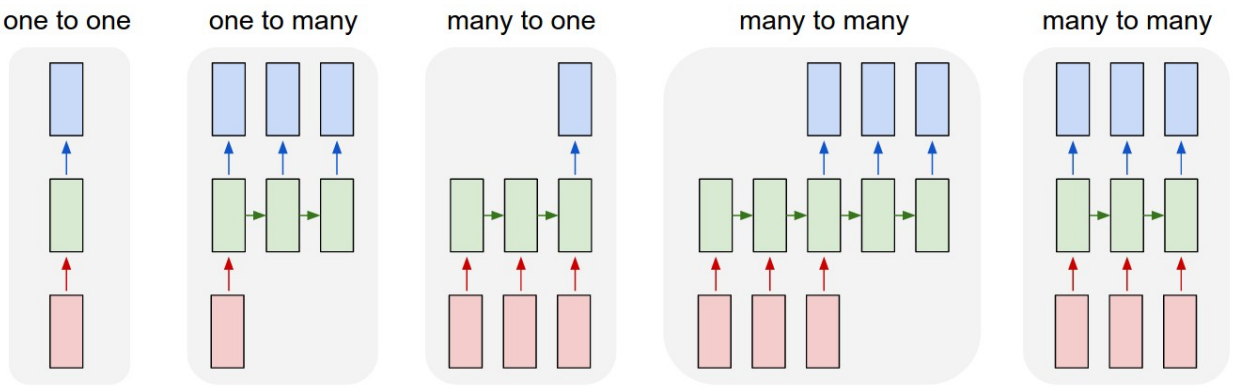


Source: https://medium.com/@udemeudofia01/basic-overview-of-convolutional-neural-network-cnn-4fcc7dbb4f17

3. *Recurrent Neural Networks (RNN)*

Used when data has a sequential structure and dependence on the past

- Common examples are LSTM and GRU's (long shot term memory models) (ways for connecting networks)
- Generate caption for image (one to many processing sequence)
    - Imaging captioning
    - Researcher: Devi Parikh
- Sentiment classification (happy or sad sentence which is a many to one process sequence)
- Pass in a set of words in one language and output those words in another language (many to many process sequencing)
- Give partial data and it is used to predict the next word in a sentence
- Have inputs and pass them into unit network model and then doing feedback loops
    - Every layer of neurons has an output and this recursively places it into the next
- Process Sequences: Can have one-to-one networks, one-to-many, many-to-one, many-to-many
- We can process a sequence of vector X by applying a recurrence formula at every time step: Ht=Fw(ht-1,x1)
    - H-1 is all represented video frames up till that point
    - X1 is input vector at some time stamp
- RNN focuses its attention at a different spatial location when generating each word
    - Recognizing the objects in each location helps for the caption
-



4. *Generator Adversarial Networks (GANS)*
   - Deep fake // creation of data→ generating new data that mimics properties of original data
   - Contains 2 parts:
     1. Generator Network
        a. generates new data for images
        b. leads to new image, video, dataset, etc.
     2. Discriminator Network
        a. Takes in the original dataset and input that's been constructed from generator network and tries to determine if it's from the original dataset or not (determine if image is fake or not)
        b. All data from generator is fake and it's trying to pass it as authentic
        c. If it incorrect classifies it, they go back to the original dataset and classify this wrong output
        d. If fake, then it's passed to the generator and the generator tries to make a more realistic image to trick the discriminator
        e. If it incorrect classifies it, they go back to the original dataset and classify this wrong output
        f. If fake, then it's passed to the generator and the generator tries to make a more realistic image to trick the discriminator