**Ruchi Shah**

**EMG 528 AI for Technology Management Final Exam 11/17/2019**

1. **Which of the following statements about machine learning is TRUE?**
2. Machine learning is the intelligence technology developed for expert machines to facilitate the learning processes of various tasks
3. **Machine learning is the intelligent technology an expert system uses to make accurate predictions**
4. Machine learning is the intelligence technology that ensure the expert computer responds accurately to its algorithm
5. Machine learning is the intelligence technology that provides computers with advanced abilities to execute processes without being specifically programmed to do so

**1. Artificial Intelligence is \_\_\_\_\_**

1. **a branch of computer science that constitutes underlying technology which enables computers to simulate human intelligence**
2. a branch of computer algorithms that facilitates an expert machine make accurate predictions
3. is a branch of computer science that renders supervision to expert machines
4. is a branch of computer science constituting underlying technology that governs expert machines

**2. Components of artificial intelligence include \_\_\_\_\_**

1. **designing, sensing, language understanding, and problem-solving**
2. upgrading, problem-solving, reasoning, and acting.
3. sensing, language understanding, problem-solving, reasoning, and acting
4. problem defining, acting, designing, and program updating

**3. In a feedforward neural network, which of the three nodes is responsible for the calculations and has no interaction with the outside world?**

1. **Input node**
2. **Hidden node**
3. **Output node**
4. Radial node

**4. Which of the six common neural network types is the most popular choice for speech recognition used in today's society?**

1. Physical
2. Feedforward
3. Modular
4. **Recurrent**

**5. Since computers can only understand binary code, what had to be introduced so that a machine could interpret the human language?**

1. **Word vectors**
2. Neural networks
3. Machine learning
4. Deep learning

**6. Why is the XOR problem exceptionally interesting to neural network researchers?**  
a) Because it can be expressed in a way that allows you to use a neural network  
b) Because it is complex binary operation that cannot be solved using neural networks  
c) Because it can be solved by a single layer perceptron  
**d) Because it is the simplest linearly inseparable problem that exists.**

**7. What is back propagation?**  
a) It is another name given to the curvy function in the perceptron  
b) It is the transmission of error back through the network to adjust the inputs  
**c) It is the transmission of error back through the network to allow weights to be adjusted so that the network can learn**  
d) None of the mentioned

**8. Why are linearly separable problems of interest of neural network researchers?**  
a) Because they are the only class of problem that network can solve successfully  
**b) Because they are the only class of problem that Perceptron can solve successfully**  
c) Because they are the only mathematical functions that are continue  
d) Because they are the only mathematical functions you can draw

**9. Which of the following is not the promise of artificial neural network?**  
**a) It can explain result**  
b) It can survive the failure of some nodes  
c) It has inherent parallelism  
d) It can handle noise

**10. Neural Networks are complex \_\_\_\_\_\_\_\_\_\_\_\_\_\_ with many parameters.**  
**a) Linear Functions**  
b) Nonlinear Functions  
c) Discrete Functions  
d) Exponential Functions

**11. A perceptron adds up all the weighted inputs it receives, and if it exceeds a certain value, it outputs a 1, otherwise it just outputs a 0.**  
**a) True**  
b) False  
c) Sometimes – it can also output intermediate values as well  
d) Can’t say

**12. Which of the following is an application of NN (Neural Network)?**  
a) Sales forecasting  
b) Data validation  
c) Risk management  
**d) All of the mentioned**

**13. An ontology**

a. provides a vocabulary for expressing knowledge

b. uses frames for hierarchical inferencing

c. is more promiscuous than perspicacious

d. represents relations, objects and properties

1. **Answer this question for either the natural language problem or the vision problem (not both):**
   1. *Vision:* You are writing a supervised learning program to identify the handwritten digits 1-9. What would your training material be? What features might your program use?
   2. ***Natural language:* You are writing a supervised learning program to translate from French to English. What would your training material be? What features might your program use?**

**A recurrent neural network (or RNN for short) is a slightly tweaked version of a neural network where the previous state of the neural network is one of the inputs to the next calculation.**

**RNNs are useful any time you want to learn patterns in data. Because human language is just one big, complicated pattern, RNNs are increasingly used in many areas of natural language processing.**

1. **Which of the following are true of convolutional neural networks (CNNs) for image analysis?**
2. **Filters in earlier layers tend to include edge detectors**
3. **Pooling layers reduce the spatial resolution of the image**
4. They have more parameters than fully connected networks with the same number of layers and the same numbers of neurons in each layer
5. A CNN can be trained for unsupervised learning tasks, whereas an ordinary neural net cannot
6. **In neural networks, nonlinear activation functions such as sigmoid, tanh, and ReLU speed up**
7. the gradient calculation in backpropagation, as compared to linear units
8. are applied only to the output units
9. **help to learn nonlinear decision boundaries**
10. always output values between 0 and 1
11. **a. Read through following Case Study on Face Detection and Answer following questions**

<https://towardsdatascience.com/face-detection-for-beginners-e58e8f21aad9>

1. List all different face detection methods and explain in 10 lines.

**Ans. Knowledge Based** – It is based on set of rules that are predefined. The rules need to be defined in detail to get the perfect answers.

**Feature based** – In the various images are scanned and the features are extracted in a structured form where the feature are broken down into small components.

**Template Matching**: This is a method where the input data is correlated with the image that is uploaded. A face model can be built by edges just by using edge detection method.

**Appearance Based:** This method is depend on a set of rule delegated to training face image to find out face models. This method is based on statistical analysis and finding out the characteristic of the feature.

1. Which are different appearance based face detection algorithms explain atleast 3 in few lines.

* **Eigenface** - based algorithm used for Face Recognition, and it is a method for efficiently representing faces using Principal Component Analysis.
* **Neutral Network** - Many detection problems like object detection, face detection, emotion detection, and face recognition, etc. have been faced successfully by Neural Networks.
* **Inductive Learning** - This approach has been used to detect faces. Algorithms like Quinlan’s C4.5 or Mitchell’s FIND-S used for this purpose.

**17b. Please read following case study and answer following questions.**

[**https://github.com/guillaume-chevalier/LSTM-Human-Activity-Recognition**](https://github.com/guillaume-chevalier/LSTM-Human-Activity-Recognition) **i. What is Long Short Term Memory LSTM? What is Sequence to Sequence.**

**Ans.** Long Short Term Memory is a kind of recurrent neural network. In RNN output from the last step is fed as input in the current step. LSTM was desgined by Hochreiter & Schmidhuber. It tackled the problem of long-term dependencies of RNN in which the RNN cannot predict the word stored in the long term memory but can give more accurate predictions from the recent information. As the gap length increases RNN does not give efficent performance. LSTM can by default retain the information for long period of time. It is used for processing, predicting and classifying on the basis of time series data.

**ii. How is LSTM applied to this use case?**

Ans. In this use case the activity is recognized from the body sensors. All the moves are recorded and a dataset is generated.

**iii. How Human Activity Recognition help in real life? Give real life example**

**Ans.** **H**uman Activity Recognition or HAR for short, is the problem of predicting what a person is doing based on a trace of their movement using sensors. Example, is the ring door bell where there is alarm everytime there is movement.

iv. Do further research and suggest latest development in Human Activity Recognition.

**Ans.** The healthcare benefits associated with regular physical activity monitoring and recognition has been considered in several research studies. Solid evidence shows that regular monitoring and recognition of physical activity can potentially assist to manage and reduce the risk of many diseases such as obesity, cardiovascular and diabetes. A few studies have been carried out in order to develop effective human activity recognition system using smartphone. However, understanding the role of each sensor embedded in the smartphone for activity recognition is essential and need to be investigated. Due to the recent outstanding performance of artificial neural networks in human activity recognition, this work aims to investigate the role of gyroscope and accelerometer sensors and its combination for automatic human activity detection, analysis and recognition using artificial neural networks. The experimental result on the publicly available dataset indicates that each of the sensors can be used for human activity recognition separately. However, accelerometer sensor data performed better than gyroscope sensor data with classification accuracy of 92%. Combining accelerometer and gyroscope performed better than when used individually with an accuracy of 95%.

**17c. Please read following case study on Architecture of AI ML end to end Pipeline.**

**https://towardsdatascience.com/architecting-a-machine-learning-pipeline-a847f094d1c7**

1. What are the steps in building AI ML pipeline.

**Ans.** 1. Data Ingestion

1. Data Preparation
2. Data Segregation
3. Model Training
4. Candidate Model Evaluation
5. Model Deployment
6. Model Scoring
7. Performance Monitoring
8. Explain each component of AIML pipeline in detail.

**Ans. 1. Data Ingestion:** Once the data is ingested, a distributed pipeline is generated which assesses the condition of the data, i.e. looks for format differences, outliers, trends, incorrect, missing, or skewed data and rectify any anomalies along the way. This step also includes the feature engineering process.

**2.Data Preparation:** Once the data is ingested, a distributed pipeline is generated which assesses the condition of the data, i.e. looks for format differences, outliers, trends, incorrect, missing, or skewed data and rectify any anomalies along the way. This step also includes the feature engineering process.

**3. Data Segregation**: The fundamental goal of the ML system is to use an accurate model based on the quality of its pattern prediction for data that it has not been trained on.

**4. Model Training:** The model training pipeline is offline only and its schedule varies depending on the criticality of the application, from every couple of hours to once a day. Apart from schedulers, the service is also time and event triggered.

**5. Candidate Model Evaluation:** This pipeline is also offline. The predictive performance of a model is evaluated by comparing predictions on the evaluation dataset with true values using a variety of metrics.

**6. Model Deployment:** The best model selected is deployed for offline (asynchronous) and online (synchronous) predictions. More than one models can be deployed at any time to enable safe transition between old and new models

**7. Model Scoring:** Model Scoring and Model Serving are two terms that are used interchangeably in the industry.

**8. Performance Monitoring:** Any ML solution requires a well-defined performance monitoring solution. An example of information that we might want to see for model serving applications includes:

• model identifier,

• deployment date/time,

• number of times the model has been served,

• average/min/max model serving times,

• distribution of features used.

• predicted vs. actual/observed results.

This metadata is calculated during the model scoring and then used for monitoring.

1. Why AI-ML pipeline architecture is so important in success of your project.

**Ans.** architecture for machine learning gives guidance for developing solution architectures where machine learning systems play a major role. Discussions on what a good architecture is, can be a senseless use of time. But input on this reference architecture is always welcome. This to make it more generally useful for different domains and different industries. Note however that the architecture as described in this section is technology agnostics. So it is aimed at getting the architecture building blocks needed to develop a solution architecture for machine learning complete.

Every architecture should be based on a strategy. For a machine learning system this means an clear answer on the question: What problem must be solved using machine learning technology? Besides a strategy principles and requirements are needed.

1. Please explain following key words and how that relates to AI ML pipeline
   1. Feature engineering: Feature engineering is the process of using domain knowledge of the data to create features that make machine learning algorithms work.
   2. Learning rate: Learning rate is a hyper-parameter that controls how much we are adjusting the weights of our network with respect the loss gradient.
   3. Regularization: Regularization in Machine Learning is an important concept and it solves the overfitting problem. It is very important to understand regularization to train a good model.
   4. Overfitting: Overfitting refers to a model that models the training data too well. Overfitting happens when a model learns the detail and noise in the training data to the extent that it negatively impacts the performance of the model on new data
   5. MSE, MAPE, F1 Score: MSE basically measures average squared error of our predictions. Mean absolute percentage error (MAPE) The mean absolute percentage error (MAPE) is a statistical measure of how accurate a forecast system is. It measures this accuracy as a percentage, and can be calculated as the average absolute percent error for each time period minus actual values divided by actual values. F1 Score. Evaluation metric for classification algorithms. F1 score combines precision and recall relative to a specific positive class -The F1 score can be interpreted as a weighted average of the precision and recall, where an F1 score reaches its best value at 1 and worst at 0.
   6. ROC Curve: ROC curve is a performance measurement for classification problem at various thresholds settings.
   7. Precision Recall: In pattern recognition, information retrieval and classification (machine learning), precision (also called positive predictive value) is the fraction of relevant instances among the retrieved instances, while recall (also known as sensitivity) is the fraction of the total amount of relevant instances that were actually retrieved.Bias Variance trade off: very few parameters then it may have high bias and low variance. On the other hand if our model has large number of parameters then it’s going to have high variance and low bias. So we need to find the right/good balance without overfitting and underfitting the data.This tradeoff in complexity is why there is a tradeoff between bias and variance. An algorithm can’t be more complex and less complex at the same time.
   8. Data augmentation: Data augmentation is a strategy that enables practitioners to significantly increase the diversity of data available for training models, without actually collecting new data.
   9. Testing: The usage of the word "testing " in relation to Machine Learning models is primarily used for testing the model performance in terms of accuracy/precision of the model. It can be noted that the word, "testing" means different for conventional software development and Machine Learning models development.
   10. Cross validation: Cross-validation is a technique for evaluating ML models by training several ML models on subsets of the available input data and evaluating them on the complementary subset of the data.
   11. Deployment and Kubernetes Kubeflow: The Kubeflow project is dedicated to making deployments of machine learning (ML) workflows on Kubernetes simple, portable and scalable. Our goal is not to recreate other services, but to provide a straightforward way to deploy best-of-breed open-source systems for ML to diverse infrastructures.

**17 d. Read the Generative Discriminative Network paper below and answer following questions:**

[**https://skymind.ai/wiki/generative-adversarial-network-gan**](https://skymind.ai/wiki/generative-adversarial-network-gan)

1. Explain Generative Adversarial Network GAN. What is generative and what is discriminative part of the GAN.

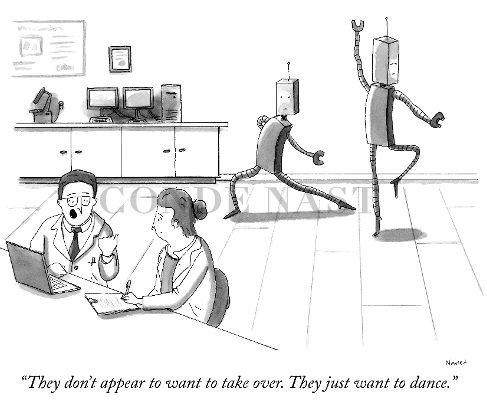
Ans. Generative Adversarial Networks (GANs) are a powerful class of neural networks that are used for unsupervised learning. It was developed and introduced by Ian J. Goodfellow in 2014. GANs are basically made up of a system of two competing neural network models which compete with each other and are able to analyze, capture and copy the variations within a dataset.

the generative network learns to map from a latent space to a data distribution of interest, while the discriminative network distinguishes candidates produced by the generator from the true data distribution.

1. Please explain Cycle GAN, Wessenstien GAN and DCGAN briefly

**Ans.** The CycleGAN is a technique that involves the automatic training of image-to-image translation models without paired examples. The models are trained in an unsupervised manner using a collection of images from the source and target domain that do not need to be related in any way.

DCGAN is one of the popular and successful network design for GAN. It mainly composes of convolution layers without max pooling or fully connected layers. It uses convolutional stride and transposed convolution for the downsampling and the upsampling.



**17e. Explain Reinforcement Learning concepts? Give your understanding of application of Reinforcement Learning.**

[**https://www.analyticsvidhya.com/blog/2017/01/introduction-to-reinforcement-learning-implementation/**](https://www.analyticsvidhya.com/blog/2017/01/introduction-to-reinforcement-learning-implementation/)

**Ans.** Reinforcement learning (RL) can be viewed as an approach which falls between supervised and unsupervised learning. It is not strictly supervised as it does not rely only on a set of labelled training data but is not unsupervised learning because we have a reward which we want our agent to maximise. The agent needs to find the “right” actions to take in different situations to achieve its overall goal.

Reinforcement learning involves no supervisor and only a reward signal is used for an agent to determine if they are doing well or not. Time is a key component in RL where the process is sequential with delayed feedback. Each action the agent makes affects the next data it receives.

1. **Answer following short questions:**

a.Which Tools/Languages should I prefer to build Deep Learning models?

**Ans.** **KNIME** – To create and productionize data science using one easy and intuitive environment, enabling every stakeholder in the data science process to be focused.

**Rapidminer** – It offers design models using a visual workflow designer or automated modeling, deploy and manage models and turn them into prescriptive actions.

**Orange** – It has a large variety of tool box and has clever data visualization tools and multidimensional data can be converted to 2D as well.

b.Why are GPUs necessary for building Deep Learning models?

**Ans.** PUs have almost 200 times more processors per chip than a CPU. For example, an Intel Xeon Platinum 8180 Processor has 28 Cores, while an NVIDIA Tesla K80 has 4,992 CUDA cores. While a CPU core is more powerful than a GPU core, the vast majority of this power goes unused by ML applications. A CPU core is designed to support an extremely broad variety of tasks (e.g., render a webpage, drive word processors and enterprise software, manage peripherals) in addition to performing computations, whereas a GPU core is optimized exclusively for data computations. Because of this singular focus, a GPU core is simpler and has a smaller die area than a CPU, allowing many more GPU cores to be crammed onto a single chip. Consequently, ML applications, which perform large numbers of computations on a vast amount of data, can see huge (i.e., 5 to 10 times) performance improvements when running on a GPU versus a CPU.

c. What is the difference between Deep Learning and Machine Learning?

**Ans.** The key difference between deep learning vs machine learning stems from the way data is presented to the system. Machine learning algorithms almost always require structured data, whereas deep learning networks rely on layers of the ANN (artificial neural networks).

d. Explain Transfer Learning with an example.  
**Ans.** In transfer learning we first train a base network on a base dataset and task, and then we repurpose the learned features, or transfer them, to a second target network to be trained on a target dataset and task. This process will tend to work if the features are general, that is, suitable to both base and target tasks, instead of being specific to the base task.

In practice, very few people train an entire Convolutional Network from scratch because it is relatively rare to have a dataset of sufficient size. Instead, it is common to pre-train a ConvNet on a very large dataset (e.g. ImageNet, which contains 1.2 million images with 1000 categories), and then use the ConvNet either as an initialization or a fixed feature extractor for the task of interest.

e. Please explain atleast 10 lessons learned from all the work and learning done during EMG 528.

**Ans.**

* How do we work on google colab
* Major difference between AI vs Machine Learning
* How image processing works
* Different techniques of processing data
* Learning about deep learning for computer vision
* Learning about how chat bots work
* Learning about real world examples where AI is used
* Learning about NLP
* Leaning how face and sound detection works
* Various tools that are used for AI and deep learning

f.Please give feedback on EMG 528. What you would like to be changed and added to this course.

Ans. The class have one or two examples to be solved so we can have a real life implication of the concepts. Maybe like a step by concept walk down.