

The Department of Computational and Data Science presents its first ever Newsletter! Through this newsletter, we hope to give you a glimpse into our little world- what we do, where our interests lie and what we hope to achieve. The Department of Computational and Data Sciences (CDS) is an interdisciplinary engineering department covering the broad research areas of computational science and engineering, and scalable computer & data systems.

Our journey began in 1970, when the Computer Centre of the institute was established as a central computing facility. In 1990, it became Supercomputer Education and Research Centre (SERC), providing computing facility to the faculty and students of the Institute. A master's degree programme M.Tech in Computational Science, one of the first of its kind in India, was started by SERC in 1999. Therefore about 16 years later, CDS was formed in December 2015 from the academic wing of SERC.

From the Chair's Desk



It is our pleasure to share the activities of the department through this newsletter. The goal of this publication is to highlight our research activities, achievements of faculty, students and project staff and new initiatives.

The main motive of this news digest is to reach out to the world outside the campus including student, faculty, corporate and any individual who is interested in computational and data sciences. Ultimately, we hope to build a highly connected, a truly inter-disciplinary society and, of course, help us to network.



Faculty Interview - with Dr. Deepak Subramani

Q1. Describe your current research.

Our current research involves developing science-based and data-driven solutions to complex engineering and environmental problems. By science-based, we mean that some sort of fundamental principles are involved. By data-driven, we mean that wherever available observational data is utilized to refine our predictions from our models of natural and engineering systems. We also imply that wherever theory is not available, or the underlying science is uncertain and/or unpredictable due to any reason, we apply techniques from data science to aid our predictions. Currently, we are focusing on three application areas: stochastic numerical ocean modeling, optimal environmental sensing, and optimal routing of ships and autonomous vehicles. Within numerical ocean modeling, we are focusing on data assimilation and use of computer vision to detect oceanic features from satellite imagery for use in our partial differential equation based ocean models. Within optimal environmental sensing, we want to answer fundamental questions about how to place sensors that help us reduce uncertainties and predict environmental states subject to cost and operational constraints of sensor deployment and maintenance. Within our routing work, we plan to extend our probabilistic PDE-based planners to assimilate new data from ships or autonomous vehicles to learn and update routes.

Q2. Why did you choose to focus on this area? Tell me about the broader significance of your research.

We are all dependent on mother nature for survival. But due to over exploitation, and unplanned use, the stress on nature is increasing. To successfully co-exist with nature, we must monitor and predict nature, which is a complex engineering task. It requires detailed knowledge about the dynamics of the natural world, rigorous uncertainty quantification, modern data collection techniques, big data and big compute systems for prediction. Such a unique combination is not only intellectually exciting, but also the skills that research students get out of doing such research will tremendously help their career. As far as possible, I want to do research that is

not only useful for advancing science and engineering, but also help students gain employable skills that after being in our group for 2 to 5 years, they can go out to the world and do great for themselves.

Q3. You joined IISc at the beginning of this year. Tell me about the lab you are setting up – how did you name it and what your future expansion plans are...

Yes, Lioined IISc on New Year's Eve! We have named our lab 'QUEST' which is an acronym for 'Quantifying Uncertainty in Engineering, Science and Technology'. There were a lot of names that I had been thinking about. I wanted a name that is inspirational, has a ring to it and, of course, reflect the kind of research that we do. If I happen to change my area after 15 years, the name should not be a liability. Hence, I decided on QUEST. U want to imbibe a culture of hard work an a no nonsense work space. I am also lucky to get good students so early. You guys are great! In the immediate future, I plan to grow my lab to have about 4 to 6 research students, 3 to 4 course students, and a few project associates and postdocs according to the funding that we can bring. I think students are my greatest ambassadors. If you all do well, then I also do well.

Q4. You did your studies at MIT. How was your student life there? How is it different from Indian institutions? MIT was a fantastic place. It is a pressure

cooker, orders of magnitude more pressure than IIT Madras where I did my UG. Course work is rigorous, and research is top class. At the same time, it was a lot of fun. My peer group was eclectic - there were water researchers who were also experts of Sanskrit language, finite element experts who were into boxing, cryptographers who were some of the best Indian and western classical musicians. Then there were pure geniuses from all over the world bringing with them knowledge of Greek philosophy, Islamic theology, American liberalism and I fondly recall some of the discussions we used to have. Of course, the kind of research that people did was pushing the boundary of human knowledge. New theories, novel applications- there was something there for everyone. I think Indian institutions are also fantastic places. There is a raw energy here.



Dr. Deepak Subramani Assistant Professor

The one thing that I see different is the way students interact with faculty.I am not saying it is good in the US and bad in India or vice versa. It is simply different. In the US, if a student does not understand a lesson, then the conclusion is that the teacher is not good. I think here, it is the reverse – if a student doesn't understand, then everyone including the students assume that they are not good.

A middle path would be good for both places. Also, in India, traditionally our student-teacher relation is sacred. It is full of reverence. We respect our teachers in thought, words and action. In the US, education is like an industry and teachers are service providers. Hence, the culture is little different. Again, I don't think one is correct and the other is wrong. The observable result of this culture difference is that in the US, research tends to be studentdriven, whereas in India, I think a major chunk of research is faculty-driven. Perhaps, with a lot of new faculty we are in flux, and new optima might emerge. We shall see.

Q5. That is nice. Why did you decide to come back to India?

When I joined for graduate work, it was clear in my mind that I will come back to India. Sometime in 2013 I think, the alumni blog of IIT Madras, called Chennai36, published an interview with the director of IITM. That interview inspired me a lot. In that he said that from his childhood he always had a sense of, for lack of a better word, patriotism. I believe the same applies to me. I cannot imagine working in a place other than India – my home. Perhaps I won't be able to achieve the heights or make as much money as I would have in the USA, but I will be happy knowing that I served my motherland and tried as best as I can, irrespective of the successes and failures that lie ahead for me.

Q6. What will your next research project be?

Since we are just starting, I think I have some bandwidth to explore, perhaps about two more application areas. So, in addition to environment related applications, I want to focus on some engineering applications, especially something with immediate strategic relevance for the country.

Q7. Let us leave research and work for a bit. Tell us what you like doing outside of research and work.

After joining IISc, I have not really had the time to do anything other than set up my lab, research, home, etc. Generally, outside work, I like to attend concerts, watch movies, YouTube series, TV shows, you know, usual stuff. I am a big fan of the Game of Thrones series.So, Monday morning 6:30 am is GoT time now – perhaps before this newsletter comes out, it will be all over for Westeros! I also like to go to old temples with some history behind it, and learn about that history. I used to play squash, and do power lifting while I was at MIT. Now, I have to find some sport here that is sustainable for my lifestyle here.

Q.8 Do you have any life advice to students?

I don't think I am qualified enough to give advice based solely on my life experience. I can only repeat the advice that I have received from my gurus – both directly and indirectly.

So, I happened to watch this speech by Harsha Bhogle on YouTube. There he said that talent can take you ahead only so much. It might give you the first step inside a door. But thereafter it is only hard work and sheer passion to perform that matters.

NEWS

He goes on to give examples of how Sachin Tendulkar practices hard – day in and day out. You take any self-made successful person in any field – arts, science, politics, business - he/she works extremely hard. My observation is that there is no exception to this. In these days of 2-day delivery, and information at fingertip, we tend to forget the importance of the need to struggle to master something.

At the same time, a journey of 1000 miles starts with step 1. Let us say, I want to build a body like say our Aquaman (or Khal Drogo, if you will), I cannot just keep that as a goal and start squatting 200 kg on Day 1. I will get dejected and stop, I can even die. I must break it down and say I will increase 2.5 kg on my squat every workout starting from zero. That is an incremental goal that is achievable. So, the advice I try to follow is to work hard and work smart towards achievable and incremental goals. Keep practicing day in and day out.

Open Day

Every year all departments across IISc showcase their activities to the student community and the general public. Thousands of visitors came to CDS for Open Day, and our hard working Open Day team, made it a grand success! Various events were organised. Each Lab presented a poster of their research. There were intriguing demonstrations of machine learning algorithms where k-means clulstering and neural network algorithms were explained. Amusing competitions like on-the-spot graph colouring, sudoku, and quizzez were held. Other interesting events included an algorithm debugging contest and talks by Dr. Deepak Subramani on "Optimal Paths of Marine Robots" and "Big Data vs Big Compute: From Weather to Election Forecasting".





Open Day Team



Open Day @ CDS



DREAM Team wins 12th IEEE SCALE Challenge, 2019

A team from the DREAM Lab won the 12th IEEE SCALE Challenge award for 2019. This is an annual award from the IEEE Technical Committee on Scalable Computing that recognizes "real-world problem solving using computing that scales", and involves a live demo and evaluation at the IEEE/ACM CCGrid conference that just concluded in Cyprus.

The winning entry was titled "**Dynamic Scaling of Video Analytics for Wide-area Tracking in Urban Spaces**". The team was led by Ph.D. candidate Aakash Khochare and included Sheshadri K. R., Shriram R., Swapnil Gandhi, Anubhav Guleria and Yogesh Simmhan. This is the first time that a team from India has won this award!

An Interview with Aakash Khochare

How would you describe your entry? What are it's application?

Smart City deployments typically have thousands to even hundreds of thousands of Surveillance cameras. Extracting actionable insights from these video feeds is compute intensive. In our TCSC SCALE Challenge 2019 entry, we explored if these video feeds can be used for societal good by creating a traffic signal "green wave" for emergency vehicles like ambulances. To this end, we designed a scalable solution that tracked the ambulance and reduced the travel time to the nearest hospital by nearly 2x in simulation.



How much effort went into the challenge? How was the support from your team? The challenge is split into two phases; an initial whitepaper submission followed by a live demonstration by shortlisted teams. Both phases took atleast 1.5 months of sustained, 'Red Bull' fuelled, effort. During the cumulative 3 months, the team was an integral part of the effort; brainstorming and implementing critical sections of our entry.

What was it like competing with teams from other renowned universities? How does it feel to be the first Indian team to win this challenge?

The other teams were equally competitive, with one team even scaling across a million cores of the Sunway TaihuLight supercomputer. However, the guidance of Prof. Yogesh Simmhan, who had won this challenge once while at USC, proved instrumental in getting us to the finish line. It's a matter of great pride to be the first Indian team to win the challenge. The victory is another example of the fact that IISc is competitive at the international stage.



Dream Lab Team in action



Aakash Khochare and Prof. Yogesh Simmhan receiving the SCALE Challenge award from George Pallis, Program Co-Chair of CCGrid 2019







Neha lyer

Swapnil Gandhi



Sangeeta Yadav

EECS Research Symposium

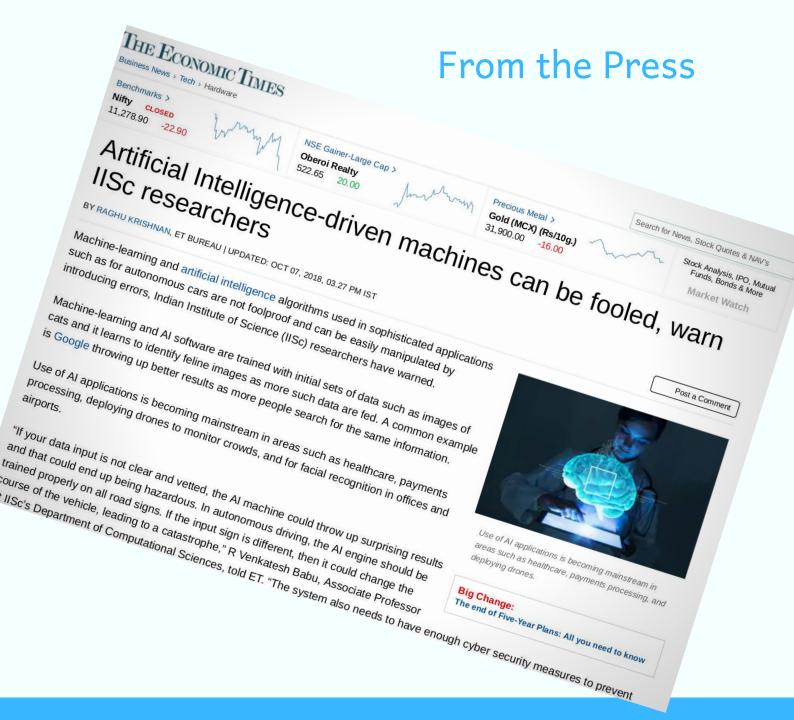
The EECS symposium is primarily a forum where Senior research students from CDS, CSA, DESE, EE, ECE and RBCCPS present their research. The symposium also features Keynote Lectures, Talks by Young Faculty and Demos from the Industry.

This year, 3 students from our department received the best poster award!

Neha Iyer for her work on Parallel hybrid smoothers in Multigrid Method for heterogeneous CPU-GPU environment

Sangeeta Yadav for her work on Robust Stabilized Finite Element Scheme for Singularly Perturbed Differential Equations using Deep Learning

Swapnil Gandhi for his work on Distributed Processing Model for Temporal Graphs



Students' Corner



Akanksha M Rajak

I joined the CDS PhD program and QUEST Lab in January 2019. I am a mathematician and I have completed my masters in mathematics from IIT Kanpur. The broad area of my research is data assimilation for regional ocean forecasting. Currently, I am working on formulating initial conditions for numerical ocean model using data assimilation techniques. Since the Indian Ocean is very vast and the data we have for it is very sparse, hence we are using different data assimilation techniques to generate the full data set. Using it as initial conditions will give better predictions.

I finished my B.Tech in Petroleum Engineering from IIT-ISM Dhanbad and currently I'm a first year M.Tech student in Computational and Data Sciences department. My work is mainly focused on ship routing. Currently I'm working on Real-time time optimal path planning of a ship, based on the environmental and oceanic conditions, as a part of my fellowship with Maersk Global Service Centre. It involves formulating the numerical solution of time optimal planning PDE and integrating observational data to make the future predictions more accurate, using data assimilation techniques.



Chennam Revanth



Raghav Sharma

I'm an M.Tech student in the department of Computational and Data Sciences. My work is focused on developing computational methods to automatically identify the Gulf Stream North Wall (GSNW) and similar currents in the ocean from satellite images.Typical methods to determine its position and boundaries require skilled human experts to do time consuming manual extraction of visualised features. These experts are essentially performing a feature extraction task, that can instead be automated to save time, guarantee objectivity, and potentially increase precision. This problem is in the domain of Image processing and artificial intelligence.

I'm a first year Ph.D. student, associated with QUEST Lab. I joined CDS, as a direct Ph.D. candidate, after having completed my B.Tech in Mechanical Engineering from NIT Warangal. I am currently working on algorithms that would allow underwater vehicles to autonomously manoeuvre in the ocean. These vehicles have sensors on them to collect data. The data needs to be collected in such a way so that the minimum number of data samples can help the vehicle perform an inference task with a fixed, satisfactory level of performance. This is the problem of sparse sensing. After obtaining the data, the vehicle must perform data assimilation and autonomously decide where/when to collect the next sample. I feel that CDS, being an interdisciplinary department, offers a lot to learn with a variety of courses. Regular seminars by experts from their fields, provide good opportunities to gain knowledge about the latest research trends.

Rohit Chowdhury

Article: Semantic Segmentation methods

Image segmentation is the process of partitioning a digital image into multiple segments that share similar attributes (like color, similar objects, etc) to simplify the representation and making it more useful for the analysis and interpretations. Segmentation defined in chit: the process of dividing something into parts or segments. In the case of an image, we have that "something" = a digital Image. For the last three decades, one of the most difficult problems in computer vision has been image segmentation. Image segmentation is different from image classification or object recognition.An object classifier will only classify objects that it have specific labels such as horse, auto, house, dog. An ideal image segmentation algorithm will also segment unknown objects, that is, objects which are new or unknown.

There are numerous applications-1. GeoSensing – For land usage

- 2. For Autonomous driving
- 2. For Autonomous driving



Fig 1

3. For Facial Recongnition

This article is dedicated to various methods of segmentation. Following are the methods with some of the key ideas.

1. Threshold Segmentation

Threshold segmentation is the simplest method of image segmentation. It is a common segmentation algorithm which directly divides the image gray scale information processing based on the gray value of different targets. Threshold can be applied locally and globally. There are effective and well established ways of thresholding an image using histogram like Otsu's algorithm. Otsu's algorithm is based on maximum interclass variance [2].The disadvantage is that it is difficult to obtain accurate results for image segmentation problems where there is no significant gray scale difference or a large overlap of the gray scale values in the image. See figure 2 for an example of thresholding based segmentatoin [1].

3. Regional Growth Segmentation

The regional growth method is a typical serial region segmentation algorithm, and its basic idea is to have similar properties of the pixels together to form a region. The method requires first selecting a seed pixel, and then merging the similar pixels around the seed pixel into the region where the seed pixel is located [1].

Seed point selection is based on some user criterion (for example, pixels in a certain gray-level range, pixels evenly spaced on a grid, etc.). The initial region begins as the exact location of these seeds. The regions are then grown from these seed points to adjacent points based on similar properties (grey level, texture and color).Since the regions are grown on the basis of the criterion, the image information itself is important. For example, if the criterion were a pixel intensity threshold value, knowledge of the histogram of the image would be of use, as one could use it to determine a suitable threshold value for the region membership criterion.

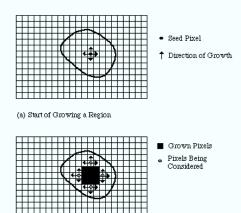
3. Segmentation based on Clustering

Clustering method is an unsupervised image segmentation method. It classifies the image into a finite number of cluster, where the number of cluster can be user defined or can be find using an algorithm. Segmentation is nothing but a classification of a pixel to its label. Clustering algorithms require a feature vector to work on, so what we do first is convert our image into a vector. This vector representation of image can be obtained by various feature extraction algorithms like SIFT, HOG ,SURF etc. In simple words these vectors are formed by using image informations like grey-scale, differential magnitude, gradient, neighbour information, etc.

The feature space clustering method is used to segment the pixels in the image space with the corresponding feature space points. According to their aggregation in the feature space, the feature space is segmented, and then they are mapped back to the original image space to get the segmentation result. K-mean clustering is one of the famous algorithm you can check the link.









4. Segmentation based on Supervised learning in CNN (deep learning)-

Deep learning have pushed the performance of computer vision tasks to new heights for a large number of highlevel problems, including image classification, object detection, video analytics. In the aspect of image segmentation, an algorithm is proposed which is more effective in this field, which is the supervised learning of a DCNN (deep convolutional neural network) for semantic image segmentation.

What is CNN ?

Convolutional Neural Networks (CNN) is one of the variants of neural networks used heavily in the field of Computer Vision. It derives its name from the type of hidden layers it consists of. The hidden layers of a CNN typically consist of convolutional layers, pooling layers, fully connected layers, and normalization layers. Here it simply means that instead of using the normal addition and activation functions, convolution and pooling functions are used respectively.

Given an image of size NxMx3, we do convolution of image with a kernel of say 3x3 size. A convolution is a mathematical operator given by:-

$$(f\ast g)(i)=\sum_{j=1}^m g(j)\cdot f(i-j+m/2)$$

CNNs are basically used to extract the feature from the image in an optimal way and after CNN we require to have fully connected layers to classify the image based on the features. The basic structure is shown in Fig 4, where each convolution layer is followed by pooling and in the end there are fully connected layers. Each layer is associated with kernels of unknown values in them, these unknowns are variables or weights for us and our goal will be obtain these weights in order to give accurate semantics. So,we make a loss function(objective function), which we needed to minimise with respect to network weights in order to match output with ground truth.

What is ground truth?

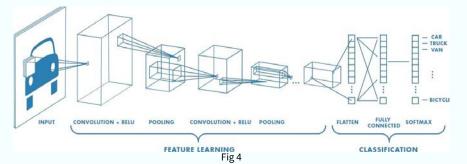
As discussed earlier it is a supervised learning algorithm so in this we have images and their correct semantically segmented images (we call them "ground truth" images).

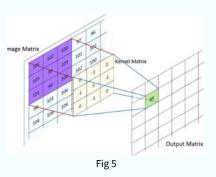
CNNs for semantic segmentation-

As we can see our CNN networks are downsampling the image as we go deep in the network. For semantic segmentation we upsample the image again to the size of the label image, then compare this output to the label image to form an objective function. The upsampling part of the network is called **decoder** and other prior part is **encoder**.

What loss function we should choose-

The most commonly used loss function for the task of image segmentation is a pixelwise cross entropy loss. This loss examines each pixel individually, comparing the class predictions(output of our network) to our one-hot encoded target vector. There are another loss functions being used like Dice loss.





The main challenges in semantic segmentation using DNNs are:

1. Reduced feature resolution, because of down-sampling of images

2. Increased training time, to scale the output features, to the resolution of input images.

3. Existence of objects at multiple scales. 4. Reduced localization accuracy due to DCNN invariance [3].

In conclusion, there are various methods available for image segmentation with varying degree of accuracy and skill. However, we are still very far from the human brain potentials and it's visual perception of things. A grand challenge of this research field is to reach the potential of human brain -to build truly artificial intelligence.

References

[1] Yuheng, S., & Hao, Y. (2017). Image Segmentation Algorithms Overview. arXiv preprint arXiv:1707.02051.

[2] Nobuyuki Otsu. (1979). A Threshold Selection Method from Gray-Level Histograms

[3] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence, 40(4):834–848, 2018.

[4] Y. Guo, Y. Liu, T. Georgiou, M.S. Lew A review of semantic segmentation using deep neural networks Int. J. Multimedia Information Retrieval, 7 (2) (2018), pp. 87-93, 10.1007/s13735-017-0141-z

Meet our alumni

An excerpt from our conversation with Mr.Srinivas Kumar

"What I enjoyed most is the fun I had with all my friends at the department. I loved how I could strike up a discussion with people on various topics. The professors here are very supportive and helpful. If I could say one thing to my juniors, it is to have good networking inside the campus and enjoy it while it lasts."



M.Tech. in Computational science ,2017. Currently,he is working as a senior associate at Capital One. His previous works include developing an OCR using deep learning and image processing. He is currently working on developing a sequel parser.

Article: Data Assimilation

Data assimilation finds extensive applications in fields like weather prediction and navigation systems. In science and engineering, we often want to predict the state of a system. A dynamic model may tell us how the state evolves with time. However, due to errors in parameters, initial conditions, boundary conditions, or in the model itself, our prediction for the state is also erroneous. These errors tend to pile up, if we only use the dynamic model for predictions! This is where data assimilation comes in. The key idea is to optimally combine (assimilate) observation data and the model to provide a better estimate of the state of the system, and thereby make more accurate predictions.



In the broadest sense, dynamic data assimilation began in the 17th century, when the planetoid Ceres disappeared behind the sun. Gauss tried to predict the position of its reappearance, using prior measurements, and in doing so, formulated the least squares method![1]

In the 1960s meteorologists started efforts to estimate the state of the global atmosphere. Data was collected from all over the world in order to make long-term predictions about the atmosphere. However, data from sparsely populated regions, like the poles and the oceans, was collected unconventionally and indirectly with the help of satellites. The question then was: how best to use information from the model (for example, a global forecast from an earlier time) and the various sources of data so as to produce an optimal estimate of the atmospheric state that is better than the model or data alone. Today, with major advances in data assimilation methods and better computational facilities, we are able to answer such questions much more accurately!

One of first successful data assimilation tools is the Kalman Filter. It estimates the internal state of a linear dynamic system from a series of noisy measurements. It is used in a wide range of applications, ranging from radar and computer vision to estimation of structural macroeconomic models. In fact, the Kalman filter was applied to navigation for the Apollo Project, which required estimates of the trajectories of manned spacecraft going to the Moon and back [2].Click here for a beautiful, short series of videos explaining how the Kalman Filter works.

Data assimilation is an intriguing field with an ever growing number of applications. Besides the classic applications in meteorology and robot control, in recent years data assimilation has been used to address challenges in life sciences, such as neuroscience, biology, biochemistry or medicine. Examples include forecast of cardio-vascular activities, estimating parameters in genetic regulatory networks, etc. In fact, with advances in AI, today, data assimilation offers researchers new frontiers to explore and develop techniques for providing more robust and accurate real time predictions.

References

[1] Lewis, J.M., Lakshmivarahan, S. and Dhall, S., 2006. Dynamic data assimilation: a least squares approach (Vol. 13). Cambridge University Press. [2]Grewal, M.S. and Andrews, A.P., 2010. Applications of Kalman filtering in aerospace 1960 to the present [historical perspectives]. IEEE Control Systems Magazine, 30(3), pp.69-78.

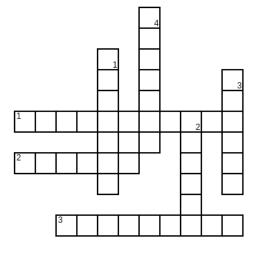


Across

- 1. the one that changes only a scalar factor when the linear transformation (Ax) is applied to it.
- 2. a node in a neural network.
- 3. facebook's object detction research platform.

Down

- A bit, contain information of an event occurring with probability half.
- 2. the one who expanded it in polynomials
- when past and future are independent given the present.
- the one who wrote the quote "Mathematics compares the most diverse phenomena and discovers the secret analogies that unite them".



For solutions - https://sites.google.com/view/solution-to-crossword-cds/home

Credits - Akanksha M Rajak, Chennam Revanth, Raghav Sharma, Rohit Chowdhury Edited by Dr. Deepak Subramani