AN INDEPENDENT STUDY BASED INTRODUCTION TO MACHINE LEARNING SYLLABUS FOR 2022 *

Lindahl, Nels
Unaffiliated Researcher
Denver, Colorado
`nelsl@nelslindahl.com`

ABSTRACT

Machine learning as a field of academic study and inquiry happens to be highly active with people publishing a tremendous amount of content at the moment. It has also graced the public mind in general with increasing popular culture references and news reporting on scientific breakthroughs. The amount of content being generated is beyond any one person to be able to categorize and consume. To help with that situation this introduction to machine learning syllabus contains an 8 part journey to set a foundation for you to begin to wade into the field of machine learning. This is an independent study based syllabus that is written without any requirement for understanding code or complex mathematics. Each section helps provide a guided subject walk with specific scholarly papers to provide depth.

Keywords Machine Learning · Syllabus · Bayesian optimization · ML Algorithms · Machine learning approaches · Neural networks · Neuroscience · ML ethics · MLOps

1 Introduction

You might remember the Substack post from week 57 titled, "How would I compose an ML syllabus?" We have now reached the point in the program where you are going to receive 8 straight Substack posts that would combine together to compose what I would provide somebody as an introduction to machine learning syllabus. We are going to begin to address the breadth and depth of the field of machine learning. Please do consider that machine learning is widely considered just a small slice of the totality of artificial intelligence research. As a spoken analogy, you could say that machine learning is just one slice of bread in the loaf that is artificial intelligence. I did seriously entertain the idea of organizing the previous 79 posts into a syllabus based format for maximum delivery efficiency. That idea gave way quickly as it would be visually and topically overwhelming and that is the opposite of how this content needs to be presented. Let's take this in the direction it was originally intended to take. To that end, let's consider the framework that back during the week 57 writing process I thought was important. My very high level introduction to the creation of a machine learning syllabus from back in week 57 on February 25, 2022, would center on 8 core topics:

- Week 80: Bayesian optimization (ML syllabus edition 1/8)
- Week 81: A machine learning literature review (ML syllabus edition 2/8)
- Week 82: ML algorithms (ML syllabus edition 3/8)
- Week 83: Machine learning Approaches (ML syllabus edition 4/8)
- Week 84: Neural networks (ML syllabus edition 5/8)
- Week 85: Neuroscience (ML syllabus edition 6/8)
- Week 86: Ethics, fairness, bias, and privacy (ML syllabus edition 7/8)
- Week 87: MLOps (ML syllabus edition 8/8)

That is what we are going to cover. At the end of the process, I'll have a first glance at an introduction to machine learning syllabus. My efforts are annotated and include some narrative compared to a pure outline based syllabus. Bringing content together that is foundational is an important part about building this collection. At this point, just describing the edge of where things are in the field of machine learning would create something that would only be current for a moment and would fade away as the technology frontier curve advances. Instead of going that route it will be better to build a strong foundation for people to consume that will support the groundwork necessary to move from introductory to advanced machine learning. Yes, you might have caught from that last sentence that at some point I'll need to write the next syllabus as a companion to this one. Stay tuned for a future advanced machine learning syllabus to go along with this introductory to machine learning edition. Enough overview has now occurred. It's time to get started...

2 Week 80: Bayesian optimization

I remember digging into Armstrong's "Principles of forecasting" book which was published back in 2001 [1]. You can get a paper copy or find it online for a lot less than the $429 dollars Springer wants for the eBook. I thought the price was a typo at first, but I don't think it actually is a typo. It's just another example of how publishers are confused about how much academic work should cost for students to be able to read. Within that weighty tome of knowledge you can find coverage of the concept of Bayesian pooling which people have used for, "Forecasting analogous time series." That bit of mathematics is always where my thoughts wander when considering Bayesian optimization. I have spent a lot of time researching machine learning and I really do believe most of the statistical foundations you would need to understand the field could be found in the book, "Principles of forecasting: A handbook for researchers and practitioners."

I do not think you should pay $429 dollars for it, but it is a wonderful book. Keep in mind that the book does not mention machine learning at all. It is from 2001 and does not really consider how forecasting tools would be extended within the field of machine learning. A lot of machine learning use cases are based on observation and the prediction of things. That is pretty much at the heart of the mathematics of forecasting. You need to understand the foundations of the statistical paradigm that Thomas Bayes introduced a couple hundred years ago in the 1700's. The outcome of that journey will be the simple aside that we are about to work toward inferring some things. Yes, at this point in the journey we are about to work on inference.

You could move directly to the point and examine Peter Frazier's 2018 "A Tutorial on Bayesian Optimization" paper [2]. You may want to extend that analysis to figure out all the connected papers. Instead of wandering off into the vast collection of papers that are connected to that one I started to wonder about a very different set of questions. You may have wondered as well if Bayesian optimization is an equation. Within the field of machine learning it is treated more like an algorithm and people typically invoke or call it from previously coded efforts. It does not appear that generally within the field of machine learning people really do the math themselves. You are going to see a whole lot of extending things that are developed as part of a package or framework. Applied Bayesian optimization is going to fall into that format of delivery and application without question.

The rest of this lecture on Bayesian optimization consists of three parts. First, 3 different videos you could watch. Second, 3 papers you could read to really dig into the subject and start to flush out your own research path. Third, an introduction to where you would find this type of effort expressed in code. Between those 3 different areas of consideration you can take your understanding of Bayesian optimization to the next level.

2.1 3 solid video explanations

- "Bayesian Optimization - Math and Algorithm Explained"
  `https://www.youtube.com/watch?v=ECNU4WIuhSE`

- "Bayesian Optimization (Bayes Opt): Easy explanation of popular hyperparameter tuning method"
  `https://www.youtube.com/watch?v=M-NTkxfd7-8`

- "Machine learning - Bayesian optimization and multi-armed bandits"
  `https://www.youtube.com/watch?v=vz3D36VXefI`

2.2 3 highly cited papers for background:

- Pelikan, M., Goldberg, D. E., & Cantú-Paz, E. (1999, July). BOA: The Bayesian optimization algorithm. In Proceedings of the genetic and evolutionary computation conference GECCO-99 (Vol. 1, pp. 525-532). [3]

- Shahriari, B., Swersky, K., Wang, Z., Adams, R. P., & De Freitas, N. (2015). Taking the human out of the loop: A review of Bayesian optimization. Proceedings of the IEEE, 104(1), 148-175. [4]
- Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical Bayesian optimization of machine learning algorithms. Advances in neural information processing systems, 25. [5]

2.3 Where would you find the code for this?

- Tensorflow
- Keras
- Scikit-learn
- A Google Colab notebook
- The base Github for the above Google Colab notebook

Closing out this lecture on Bayesian optimization has to end with a general bit of caution about the mathematics of machine learning. A lot of very complex mathematics including statistical devices are available to you within the machine learning space. Working toward a solid general understanding of what the underlying methods (especially the statistical methods) are doing is really important as a foundation for your future work. It is easy to allow the software to pick up the slack and to report outputs. Moving purely toward this type of effort allows the potential for problematic internal breakdowns of the mathematics to occur. You may very well get the outcome you wanted, but it is not explainable or repeatable in any way shape or form. Yes, I'm willing to accept that the majority of people working within the machine learning space could not take a step back and express their work in a pure mathematical way by abstracting away the code to a pure equation based form. That type of pure mathematical explanation by equation is not generally required in papers or read outs. Most of the time it comes down to the simple truth of working in production.

3 Week 81: A machine learning literature review

You can find a lot of quality explanations of the differences between the various flavors of machine learning. This second lecture in the introduction to ML syllabus series should open with a series of the best literature reviews I could find and pull together to share. That will be the second part of this lecture. The third part will cover the intersection of programming languages. Some rather high quality textbooks and manuscripts exist within the field of machine learning. You can even find ones for free on GitHub and other places. Instead of starting with the obvious way to go by digging into some weighty tomes. I'm going to spend some time sharing readouts of some of the most highly cited machine learning papers. For a lot of people jumping into the field they are working on something in a different field of study and find a use case or a business related adventure that could benefit from machine learning. Typically at this point they are going to start digging into software and can get going very rapidly. That part of the journey requires no real deep dive into the relevant literature. It's great that people can just jump in and find machine learning accessible. However, (you knew that was coming) the next phase in the journey is when people start wondering about the why and how of what is happening or they dig deep enough that they may want to know about the foundations of the technology or techniques they are using. At that point, depending on what is being done people will see a massive number of papers published and shared online. The vast majority are available to freely download and read.

3.1 Part 1: Highly cited machine learning papers

Within this section I'm going to try to build out a collection of 10 things you could read to start getting a sense of what papers within the machine learning space are highly cited. That is not a measure of readability or how solid of a literature review for machine learning they provide. You will find that most of them do not have really lengthy literature sections. The authors make the citations they need to make for related work and jump into the main subject pretty quickly. I'm guessing that is a key part of why they are highly cited publications. To begin with; from what I can tell, the most highly cited and widely shared paper of all time in the machine learning or deep learning space has over 125,285 citations that Google Scholar is aware of and can index. That is the first paper in the list below.

1. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778). This paper is cited a ton of times and has a pretty solid references section. If you read it after seeing the link above, then you would run into a bit of introduction on deep convolutional neural networks and then it would jump into some related work sections on residual representations, shortcut connections, and finally deep residual learning. While this paper is cited well over one hundred thousand times it is not designed to be an introduction to machine learning. It's

12 pages and it provides a solid explanation of using deep residual learning for doing image recognition. To that end, this paper is highly on point and easy to read which is probably why so many people have cited it from 2016 to now. [6]

2. Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255-260. Within the start of this review you are going to get a lot more of an introduction to what machine learning involves and I'm not surprised this work is highly cited. [7]

3. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. Nature 521, 436–444 (2015). This one is a very readable paper. It was certainly written to be widely read and is very consumable. It has 103 citations as well which is an intense number. [8]

4. Ioffe, S., & Szegedy, C. (2015, June). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In International conference on machine learning (pp. 448-456). PMLR. [9]

5. Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems, 28. [10]

6. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30. [11]

7. Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473. [12]

8. Mnih, V., Kavukcuoglu, K., Silver, D. et al. Human-level control through deep reinforcement learning. Nature 518, 529–533 (2015). [13]

9. Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based learning applied to document recognition," in Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, Nov. 1998, doi: 10.1109/5.726791. [14]

10. Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980. [15]

During part one of this lecture I covered 10 different machine learning papers that are highly cited. My top 10 list might very well not be your top 10 list. If you have a different one, then feel free to share it as well. I'm open to criticism and alternative methods. You can work the paper and reference journey to start to get a solid understanding of machine learning. That is one way to go about getting an introduction to the field. That method involves reading key pieces of literature and as you see footnotes and references that would help fill in your knowledge you take the time to work your way backward from anchor to anchor completing a highly personalized literature review. For academics or people highly focused on a special area within the academic space this is a tried and true method for learning. People are doing it all the time in business and in graduate schools all over the world. Another method exists as well and we will explore that more next.

3.2   Part 2: General literature reviews, text books, and manuscripts about machine learning

Sometimes you just want to have all the content packaged up and provided to you as a single serving introduction to machine learning. I'm aware that within this lecture I did not elect to take that single serving path. This field of study is large enough and includes a diverse enough set of knowledge that I think you need to approach it in a variety of different ways based on your specific learning needs. To that end I broke my machine learning literature review into two distinct parts. This second part is about where you could pick up one source and get started, but hopefully it won't be the final destination in the lifeline learning journey that is understanding the ever changing field of machine learning. For those of you who have been reading this series for sometime you know that my go to introductory text is from the field of artificial intelligence and would be Stuart Russel and Peter Norvig's classic "Artificial Intelligence: A Modern Approach" which is in its 4th edition based on the Berkeley website [16]. I have the 3rd edition on my bookshelf that I picked up on eBay. The 4th edition has a whole section devoted to machine learning including: learning from examples, probabilistic models, deep learning, and reinforcement learning. That is certainly a popular place to start for people who are starting to dig into machine learning and probably more importantly want a solid foundation in artificial intelligence as well.

- You could go with a classic from 1997 and start with a book literally called "Machine Learning" by Tom Mitchell. Mitchell, T. M. (1997). Machine learning. New York: McGraw-hill. [17]

- Maybe you were looking for something a little newer than 1997. You could jump over to the freely available Deep Learning book by Ian Goodfellow, Yoshua Bengio, and Aaron Courville that was published back in 2016. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press. [18]

A lot more books exist that could help give you an introduction to machine learning, but I'm going to close out with the three that I happen to like the best. That does not mean they are the only way to go about learning machine learning.

3.3 Part 3: All the code based introduction to machine learning efforts

I'm going to let my TensorFlow bias run wild here for a moment and say that on my bookshelf right now are a few different works from Valliappa Lakshmanan. Within the TensorFlow community you will find a ton of well written and interesting sets of videos, courses, and other content that will help you dig into the field of machine learning. Outside of the TensorFlow content and the myriad of works by Lak I have a few other books on my bookshelf worth mentioning. I'm not sure why they are all published by O'Reilly, but that appears to be a theme of what made it to my bookshelf in terms of coding books. I know buying and subsequently keeping physical books is something that I do when I'm first learning something. I find it comforting to see them sitting next to me on my bookshelf in my office.

- Grus, J. (2019). Data science from scratch: first principles with python. O'Reilly Media. [19]
- Hope, T., Resheff, Y. S., & Lieder, I. (2017). Learning tensorflow: A guide to building deep learning systems. O'Reilly Media. [20]
- Graesser, L., & Keng, W. L. (2019). Foundations of deep reinforcement learning: theory and practice in Python. Addison-Wesley Professional. [21]

3.4 Part 4: Super brief conclusion

Within this brief introduction to machine learning literature review we covered the top 10 articles I think you should start out reading and then we dug into the top 3 textbooks that stood out to me. During the first lecture you might also remember that in terms of forecasting and statistics another book was recommended. It had nothing to do with machine learning, but it's a solid foundational textbook for people interested in understanding the statistics of forecasting.

- Armstrong, J. S. (Ed.). (2001). Principles of forecasting: a handbook for researchers and practitioners (Vol. 30). Boston, MA: Kluwer Academic. [1]

Other introduction to statistical methods books exist and one of them might be right for you if you need to brush up on some of the mathematics that you will encounter within the machine learning space. Beyond that, hopefully this lecture has given you a brief introduction to the treasure trove of literature available to give you an introduction to machine learning.

4 Week 82: ML algorithms

Welcome to the lecture on ML algorithms. This topic was held until the 3rd installment of this series to allow a foundation for the concept of machine learning to develop. At some point, you are going to want to operationalize your knowledge of machine learning to do some things. For the vast majority of you one of these ML algorithms will be that something. Please take a step back and consider this very real scenario. Within the general scientific community getting different results every time you run the same experiment makes publishing difficult. That does not stop authors in the ML space. Replication and the process of verifying scientific results is often difficult or impossible without similar setups and the same data sets. Within the machine learning space where a variety of different ML algorithms exist that is a very normal outcome. Researchers certainly seem to have gotten very used to getting a variety of results. I'm not talking about using post theory science to publish based on allowing the findings to build knowledge instead of the other way around. You may very well get slightly different results every time one of these ML algorithms is invoked. You have been warned. Now let the adventure begin. One of the few Tweets that really made me think about the quality of ML research papers and the research patterns impacting quality was from Yaroslav Bulatov who works on the PyTorch team back on January 22, 2022. That tweet referenced a paper on ArXiv called, "Descending through a Crowded Valley — Benchmarking Deep Learning Optimizers," from 2021 [22]. That paper digs into the state of things where hundreds of optimization methods exist. It pulls together a really impressive list. The list itself was striking just in the volume of options available. My next thought was about just how many people are contributing to this highly overcrowded field of machine learning. That paper about deep learning optimizers covered a lot of ground and would be a good place to start digging around. We are going to approach this a little differently based on a look at the most common ones.

Here are some (10) very common ML algorithms (this is not intended to be an exhaustive list):

1. XGBoost

2. Naive Bayes algorithm
3. Linear regression
4. Logistic regression
5. Decision tree
6. Support Vector Machine (SVM) algorithm
7. K-nearest neighbors (KNN) algorithm
8. K-means
9. Random forest algorithm
10. Diffusion

I'm going to talk about each of these algorithms briefly or this would be a very long lecture. We certainly could go all hands and spend several hours all in together in a state of irregular operations covering these topics, but that is not going to happen today. To make this a more detailed syllabus version of the lecture I'm going to include a few references to relevant papers you can get access to and read after each general introduction. My selected papers might not be the key paper or the most cited. Feel free to make suggestions if you feel a paper better represents the algorithm. I'm open to suggestions.

XGBoost - Some people would argue with a great deal of passion that we could probably be one and done after introducing this ML algorithm. You can freely download the package for this one on GitHub. It has over 20,000 stars on GitHub and has been forked over 8,000 times. People really seem to like this one and have used it to win competitions and generally get great results. Seriously, you will find references to XGBoost all over these days. It has gained a ton of attention and popularity. Not exactly to the level of being a pop culture reference, but within the machine learning community it is well known. The package is based on gradient boosting and provides parallel tree boating (GBDT, GBM). This package generally creates a series of models that boost the trees and help create overfitting in sequential efforts. You can read a paper from 2016 about it on arXiv called, "XGBoost: A Scalable Tree Boosting System". The bottom line on this one is that you get a lot of benefits from gradient boosting built into a software package that can get you moving quickly toward your goal of success.

- Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794). [23]
- Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., & Chen, K. (2015). Xgboost: extreme gradient boosting. R package version 0.4-2, 1(4), 1-4. [24]

Naive Bayes algorithm - You knew I would have to have something Bayes related near the top of this list. This one is a type of classifier that helps evaluate the probability or relationship between classes. One of the classes with the highest probability will be considered the most likely class. It also assumes that those features are independent. I found a paper on this one that was cited about 4,146 times called, "An empirical study of the naive Bayes classifier".

- Rish, I. (2001, August). An empirical study of the naive Bayes classifier. In IJCAI 2001 workshop on empirical methods in artificial intelligence (Vol. 3, No. 22, pp. 41-46). [25]

Linear regression - This is the most basic algorithm and statistical technique in use here where based on a line (linear) a relationship can be charted for prediction between two things. A lot of the graphics you will see where a lot of content is mapped on a chart with a line dividing the general middle of the distribution would potentially be using some form of linear regression.

- Forkuor, G., Hounkpatin, O. K., Welp, G., & Thiel, M. (2017). High resolution mapping of soil properties using remote sensing variables in south-western Burkina Faso: a comparison of machine learning and multiple linear regression models. PloS one, 12(1), e0170478. [26]
- Maulud, D., & Abdulazeez, A. M. (2020). A review on linear regression comprehensive in machine learning. Journal of Applied Science and Technology Trends, 1(4), 140-147. [27]

Logistic regression - This type of statistical model allows an algorithmic analysis of the probability of success or failure. You could model other binary type questions. The good folks over at IBM have an entire set of pages set up to run through how logistic regression could be a tool to help with decision making. This model is everywhere in simple analysis of things when people are trying to work toward a single decision.

- Christodoulou, E., Ma, J., Collins, G. S., Steyerberg, E. W., Verbakel, J. Y., & Van Calster, B. (2019). A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. Journal of clinical epidemiology, 110, 12-22. [28]
- Dreiseitl, S., & Ohno-Machado, L. (2002). Logistic regression and artificial neural network classification models: a methodology review. Journal of biomedical informatics, 35(5-6), 352-359. [29]

Decision tree - Imagine diagramming decisions and coming to a fork where you have to decide to go one way or the other. That is how decision trees work based on inputs and corresponding outputs. Normally you will have a bunch of interconnected forks in the road and together they form up a decision tree. A lot of really great explanations of this exist online. One of my favorite ones is from Towards Data Science and was published way back in 2017.

- Dietterich, T. G., & Kong, E. B. (1995). Machine learning bias, statistical bias, and statistical variance of decision tree algorithms (pp. 0-13). Technical report, Department of Computer Science, Oregon State University. [30]

Support Vector Machine (SVM) algorithm - You are going to need to imagine graphing out a bunch of data points then trying to come up with a line that separates them with a maximum margin. A solid explanation of this can be found within a Towards Data Science article from 2018.

- Noble, W. S. (2006). What is a support vector machine?. Nature biotechnology, 24(12), 1565-1567. [31]
- Wang, L. (Ed.). (2005). Support vector machines: theory and applications (Vol. 177). Springer Science & Business Media. [32]
- Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J., & Scholkopf, B. (1998). Support vector machines. IEEE Intelligent Systems and their applications, 13(4), 18-28. [33]

K-nearest neighbors (KNN) algorithm - Our friends over at IBM are sharing all sorts of knowledge online including a bit about the KNN algorithm. Apparently, the best commentary explaining this one comes from Sebastian Raschka back in the fall of 2018. This one is pretty much what you would expect from a technique that looks at distance between neighboring points.

- Peterson, L. E. (2009). K-nearest neighbor. Scholarpedia, 4(2), 1883. [34]
- Zhang, M. L., & Zhou, Z. H. (2005, July). A k-nearest neighbor based algorithm for multi-label classification. In 2005 IEEE international conference on granular computing (Vol. 2, pp. 718-721). IEEE. [35]

K-means - Some algorithms work to evaluate clusters and K-means is one of those. You can use this to try to help classify unlabeled data into clusters which can be helpful.

- Sinaga, K. P., & Yang, M. S. (2020). Unsupervised K-means clustering algorithm. IEEE access, 8, 80716-80727. [36]

Random forest algorithm - Most of the jokes that have been told within the machine learning space often relate to decision trees. The field is not full of a lot of jokes, but trees falling in a random forest are often included in that branch. People really liked the random forest algorithm for a time. You can imagine that a bunch of trees are created to engage in the prediction of classification. The random tree in the forest with the best classification production becomes the winner. This is great as it could find something that was noval or unexpected result based on the randomness.

- Biau, G., & Scornet, E. (2016). A random forest guided tour. Test, 25(2), 197-227. [37]

Diffusion - Previously I covered diffusion back in week 79 to try to figure out why it is becoming so popular. It is in no way as popular as XGBoost, but it has been gaining popularity. Over in the field of thermodynamics you could study gas molecules. Maybe you want to learn about how those gas molecules would diffuse from a high density to a low density area and you would also want to know how those gas molecules would reverse course. That is the basic theoretical part of the equation you need to absorb at the moment. Within the field of machine learning people have been building models that learn how based on degree of noise to diffuse the data and then reverse that process. That is basically the diffusion process in a nutshell. You can imagine that the cost to do this is computationally expensive.

- Wei, Q., Jiang, Y., & Chen, J. Z. (2018). Machine-learning solver for modified diffusion equations. Physical Review E, 98(5), 053304. [38]

- Dhariwal, P., & Nichol, A. (2021). Diffusion models beat gans on image synthesis. Advances in Neural Information Processing Systems, 34, 8780-8794. [39]

Wrapping this lecture up should be pretty straightforward. Feel free to dig into some of those papers if anything grabbed your attention this week. A lot of algorithms exist in the machine learning space. I tried to grab algorithms that are timeless and will always be relevant when considering where machine learning as a field is going.

5 Week 83: Machine learning approaches

During the last lecture we jumped in and looked at 10 machine learning algorithms. This week the content contained within this lecture will cover from a machine learning perspective reinforcement learning and 3 types of supervised learning. Those types of supervised learning will include the general use case of supervised learning, unsupervised learning, and the super interesting semi-supervised learning. Like the model for consideration used in the last lecture I'll cover the topics in general and provide links to papers covering the topic to allow people looking for a higher degree of depth to dive deeper into academic papers to achieve that goal. My general preference here is to find academic papers that are both readable and are generally available for you to actually read with very low friction. Within the machine learning and artificial intelligence space a lot of papers are generally available and that is great for literature reviews and generally for scholarly work and practitioners working to implement the technology. My perspective is a mix between those two worlds which could be defined as a pracademic view of things. All right; here we go.

Reinforcement learning - Welcome to the world of machine learning. This is probably the first approach you are going to learn about in your journey. That's right, it's time to consider for a brief moment the world of reinforcement learning. You are probably going to need to start to create some intelligent agents and you will want to figure out how to maximize the reward those agents could get. One method of achieving that result is called reinforcement learning. A lot of really great tutorials exist trying to explain this concept and one that I enjoyed was from Towards Data Science way back in 2018. The nuts and bolts of this one involve trial and error with an intelligent agent trying to learn from mistakes using a maximization of reward function to avoid going down paths that don't offer greater reward. The key takeaway here is that during the course of executing a model or algorithm a maximization function based on reward has to be in place to literally reinforce maximization during learning. I'm sharing references and links to 4 academic papers about this topic to help you dig into reinforcement learning with a bit of depth if you feel so inclined.

- Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: A survey. Journal of artificial intelligence research, 4, 237-285. [40]
- Sutton, R. S., & Barto, A. G. (1998). Introduction to reinforcement learning. [41]
- Szepesvári, C. (2010). Algorithms for reinforcement learning. Synthesis lectures on artificial intelligence and machine learning, 4(1), 1-103. [42]
- Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., & Riedmiller, M. (2013). Playing atari with deep reinforcement learning. arXiv preprint arXiv:1312.5602. [43]

Supervised learning - You knew it would only be a matter of time before we went out to some content from our friends over at IBM. They note that within a world where you have some labeled datasets and are training an algorithm to engage in classification or perhaps regression, but probably classification. In some ways the supervised element here is the labeling and guiding of the classification. Outside of somebody or a lot of people sitting and labeling training data the supervision is not from somebody outright sitting and watching the machine learning model run step by step. Some ethical considerations need to be taken into account at this point. A lot of people have worked to engage in data labeling. A ton of services exist to help bring people together to help do this type of work. Back in 2018 Maximilian Gahntz published a piece in Towards Data Science that talked about the invisible workers that are doing all that labeling in large curated datasets. Within the world of supervised learning being able to get high quality labeled data really impacts the ability to make solid models. It's our ethical duty as researchers to consider what that work involves and who is doing that work. Another article in the MIT Technology Review back in 2020 covered the idea of how gig workers are powering a lot of this labeling. The first academic article linked below with Saiph Savage as a co-author will cover the same topic and you should consider giving it a read to better understand how machine learning is built from dataset to model. After that article, the next two are general academic articles about predicting good probabilities and empirical comparisons to help ground your understanding of supervised learning.

- Hara, K., Adams, A., Milland, K., Savage, S., Callison-Burch, C., & Bigham, J. P. (2018, April). A data-driven analysis of workers' earnings on Amazon Mechanical Turk. In Proceedings of the 2018 CHI conference on human factors in computing systems (pp. 1-14). [44]

- Niculescu-Mizil, A., & Caruana, R. (2005, August). Predicting good probabilities with supervised learning. In Proceedings of the 22nd international conference on Machine learning (pp. 625-632). [45]
- Caruana, R., & Niculescu-Mizil, A. (2006, June). An empirical comparison of supervised learning algorithms. In Proceedings of the 23rd international conference on Machine learning (pp. 161-168). [46]

Unsupervised learning   - It's a good thing that you were paying very close attention to the explanation of supervised learning above. Imagine that the humans or in some cases the vast collectives of humans labeling training sets just stopped doing that. Within the unsupervised learning world the classification within the machine learning problem space is going to be handed differently. Labeling and the creation of classification has to be a part of the modeling methodology. This topic always makes me think of the wonderful time capsule of a technology show about startups called Silicon Valley (2014 to 2019) that was broadcast by HBO. They had an algorithm explained at one point as being able to principally identify food as hot dog or not hot dog. That's it the model only could do the one task. It was not capable of correctly identifying all food as that is a really complex task. Trying to use unsupervised learning for example, based on tags and other information identifying different types of food in photographs is something that people have certainly done with unsupervised learning approaches. I'm only sharing one paper about this approach and its from 2001.

- Hofmann, T. (2001). Unsupervised learning by probabilistic latent semantic analysis. Machine learning, 42(1), 177-196. [47]

Semi-supervised learning   - All 3 of these different types of learning supervised, unsupervised, and semi-supervised are related. They are different methods of attacking a problem space related to learning as part of the border landscape of machine learning. You can imagine that people wanted to try to create a hybrid model when a limited set of labeled data is used to help begin the modeling process. That is the essence of the process of building out a semi-supervised learning approach. You could read more about that over on Towards Data Science. I'm sharing 3 different academic papers related to this topic that cover a literature review, a book about it, and the more advanced topic of pseudo labeling.

- Zhu, X. J. (2005). Semi-supervised learning literature survey. [48]
- Chapelle, O., Scholkopf, B., & Zien, A. (2009). Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews]. IEEE Transactions on Neural Networks, 20(3), 542-542. [49]
- Lee, D. H. (2013, June). Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In Workshop on challenges in representation learning, ICML (Vol. 3, No. 2, p. 896). [50]

Conclusion   - This lecture covered reinforcement learning and 3 types of supervised learning. You could spend a lot of time digging into academic articles and books related to these topics. Generally, I believe you will start to want to look at use cases and direct your attention to highly specific examples of applied machine learning at this point. Fortunately, a lot of those papers exist and you won't be disappointed.

6   Week 84: Neural networks

You may find in the literature that this topic of neural networks is sometimes called the zoo or more specifically, "the neural network zoo." Corresponding to the articles that make this reference is a wonderful included graphic that shows a ton of different neural networks and can really give you a sense of how they work at the most fundamental level. Two papers that make this reference and include that wonderful graphic are from researchers at the Asimov institute that had papers published in 2016 and 2022. Both of those papers are great places to start learning about neural networks.

- Van Veen, F., & Leijnen, S. (2016). The neural network zoo. The Asimov Institute. [51]
- Leijnen, S., & Veen, F. V. (2020). The neural network zoo. Multidisciplinary Digital Publishing Institute Proceedings, 47(1), 9. [52]

That brief introduction aside. We are now going to focus on specific types of neural networks and next week our focus will shift to the topic of neuroscience. I have separated the two topics on purpose. Briefly, I had considered trying to combine the two topics as one set of content, but I think it would have become unwieldy in terms of trying to present a distinct point of view on both topics. Digging into neural networks is really about dig into deep learning and trying to understand it as a subfield of machine learning. Keep in mind that while machine learning is exciting it's just a small part of the broader grouping of artificial intelligence as a field of study. I'm going to provide a brief introduction and some links to scholarly articles for 9 types of neural networks that you might run into. This list is in no way

comprehensive and is built and ordered based on my interests as a researcher. A lot of speciality models and methods exist. One of them could end up displacing something on the list if it proves highly effective. I'm open to suggestions of course for different models or even orders of explanation.

1. Artificial Neural Networks (ANN)

2. Simulated Neural Networks (SNN)

3. Recurrent Neural Networks (RNN)

4. Generative Adversarial Network (GAN)

5. Convolutional Neural Network (CNN)

6. Deep Belief Networks (DBN)

7. Self Organizing Neural Network (SONN)

8. Deeply Quantized Neural Networks (DQNN)

9. Modular Neural Network (MNN)

Artificial Neural Networks (ANN)   - This is the model that is generally shortened to just neural networks and it is a very literal title. An ANN is really an attempt or more accurately a computational model designed to either mimic or create a neural network akin to what is used within a biological brain using hardware or software. You can assume this model to be fundamental to any consideration of neural networks, but you are going to quickly want to dig into other more targeted models based on your specific use case. What you are trying to accomplish will certainly help you focus on a model or method that best meets the needs of that course of action. However, in the abstract people will consider how to build ANNs and what they could be used for as the technology progresses.

- Jain, A. K., Mao, J., & Mohiuddin, K. M. (1996). Artificial neural networks: A tutorial. Computer, 29(3), 31-44. [53]

- Hassoun, M. H. (1995). Fundamentals of artificial neural networks. MIT press. [54]

Simulated Neural Networks (SNN)   - As you work along your journey in the deep learning space and really start to dig into neural networks you will run into those ANNs and very quickly a subset of machine learning adjacent to that type of model called the simulated neural networks. Creating a neural network that truly mimics the depth and capacity of the brain is something to strive for right now and with that constraint it makes sense that work is being done to simulate the best possible representation we can achieve currently or a very special use case that limits the simulation. Using models that generate a simulation based on some complex sets of mathematics, these SNNs are being created to challenge certain use cases. One of the papers shared below is associated with figuring out the shelf life of processed cheese for example.

- Kudela, P., Franaszczuk, P. J., & Bergey, G. K. (2003). Changing excitation and inhibition in simulated neural networks: effects on induced bursting behavior. Biological cybernetics, 88(4), 276-285. [55]

- Goyal, S., & Goyal, G. K. (2012). Application of simulated neural networks as non-linear modular modeling method for predicting shelf life of processed cheese. Jurnal Intelek, 7(2), 48-54. [56]

Recurrent Neural Networks (RNN)   - At some point you will want to move from simulating and modeling to accomplishing the hard work of applied machine learning for a specific use case. One of the models you will see being used actively are variations and direct implementations of recurrent neural networks. Within this type of model patterns are going to be identified within the data and the modeling will be based on those patterns to engage in a prediction of the most likely next scenario. This is a useful approach for speech recognition or handwriting analysis. You probably have run into an RNN at some point today with your smartphone or a connected home speaker. A lot of very interesting applied use cases exist for RNNs.

- Lipton, Z. C., Berkowitz, J., & Elkan, C. (2015). A critical review of recurrent neural networks for sequence learning. arXiv preprint arXiv:1506.00019. [57]

- Yin, C., Zhu, Y., Fei, J., & He, X. (2017). A deep learning approach for intrusion detection using recurrent neural networks. Ieee Access, 5, 21954-21961. [58]

Generative Adversarial Network (GAN)   - For me personally, this is where things get interesting. Instead of looking at one neural network this GAN model creates the possibility of gamification or more to the point direct competition between models in an adversarial way. Two generative models or potentially more can be compared to figure out an optimal approach. I think this is a very interesting methodology and one that could yield very interesting futur results. You can read a lot about this and see the early code published about 8 years ago from Ian Goodfellow over on GitHub.

- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. Advances in neural information processing systems, 27. [59]
- Yi, X., Walia, E., & Babyn, P. (2019). Generative adversarial network in medical imaging: A review. Medical image analysis, 58, 101552. [60]
- Aggarwal, A., Mittal, M., & Battineni, G. (2021). Generative adversarial network: An overview of theory and applications. International Journal of Information Management Data Insights, 1(1), 100004. [61]

Convolutional Neural Network (CNN)   - You will run into use cases where you want to dig into visual imagery and that is where CNNs will probably pop up very quickly. You are building a model or algorithm that based on weights and biases can evaluate a series of images or potentially other content. The process of how layers are made and what exactly fuels a CNN is a very interesting process of abstraction.

- Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017, August). Understanding of a convolutional neural network. In 2017 international conference on engineering and technology (ICET) (pp. 1-6). Ieee. [62]
- O'Shea, K., & Nash, R. (2015). An introduction to convolutional neural networks. arXiv preprint arXiv:1511.08458. [63]

Deep Belief Networks (DBN)   - You may have run into this one in the news recently with all the coverage related to drug discovery. DBNs are frequently described as graphical in nature and generative. The reason it works for something as influential and interesting as drug discovery is that you can produce all the possible values for potential new drugs in a use case and evaluate those results. This is an area where I think the things being produced will be extremely beneficial assuming the methodology is used in positive ways.

- Salakhutdinov, R., & Murray, I. (2008, July). On the quantitative analysis of deep belief networks. In Proceedings of the 25th international conference on Machine learning (pp. 872-879). [64]
- Hinton, G. E. (2009). Deep belief networks. Scholarpedia, 4(5), 5947. [65]

Self Organizing Neural Network (SONN)   - Imagine a neural network model based on feature maps or Kohonen maps that is unsupervised and self-organizing. Within that explanation you are going to get a self organizing neural network model. This could be used for adaptive pattern recognition or just regular pattern recognition. The two references shared below will spell out how this works in more detail if you are interested.

- Carpenter, G. A., & Grossberg, S. (1988). The ART of adaptive pattern recognition by a self-organizing neural network. Computer, 21(3), 77-88. [66]
- Carpenter, G. A., & Grossberg, S. (Eds.). (1991). Pattern recognition by self-organizing neural networks. MIT Press. [66]

Deeply Quantized Neural Networks (DQNN)   - Within a neural network model when you are creating weights you could elect to use only very small ones from 1 to 8 bits and to that end you would be on your way to a deeply quantized neural network. Development tools exist for this type of effort like the Google team's qKeras and Larq. Getting open access to papers on this topic is a little harder than some of the others, but you can pretty quickly get to the code on how to implement this type of neural network.

- Dogaru, R., & Dogaru, I. (2021). LB-CNN: An Open Source Framework for Fast Training of Light Binary Convolutional Neural Networks using Chainer and Cupy. arXiv preprint arXiv:2106.15350. [67]

Modular Neural Network (MNN)   - Within this model you are going to want to create independent neural networks and moderate them. Within this framework each independent neural network is a module of the whole. This one always makes me think of building blocks for some reason, but that is a simplistic representation given the ability for moderation required to make this work.

- Devin, C., Gupta, A., Darrell, T., Abbeel, P., & Levine, S. (2017, May). Learning modular neural network policies for multi-task and multi-robot transfer. In 2017 IEEE international conference on robotics and automation (ICRA) (pp. 2169-2176). IEEE. [68]
- Happel, B. L., & Murre, J. M. (1994). Design and evolution of modular neural network architectures. Neural networks, 7(6-7), 985-1004. [69]

Conclusion - This is an intense way of starting to dig into neural networks and you will very quickly see that the use cases outside of pure machine learning or artificial intelligence are driving this field forward. A lot of these use cases are within the medical field or health care in general and are super interesting and somewhat related to neuroscience. That is where the next lecture will head in this series. Discussion will move from specific types of neural networks and the research associated with them to the broader topic of neuroscience and how it relates to machine learning.

7 Week 85: Neuroscience

Neuroscience is a complex topic to dig into in general. Within the context of machine learning it gets even more interesting. Understanding that context of complexity it will make sense here to focus on 5 scholarly articles that could help provide a solid context here for the relationship between neuroscience and machine learning.

1. Savage, N. (2019). How AI and neuroscience drive each other forwards. Nature, 571(7766), S15-S15. [70]
2. Richards, B. A., Lillicrap, T. P., Beaudoin, P., Bengio, Y., Bogacz, R., Christensen, A., ... & Kording, K. P. (2019). A deep learning framework for neuroscience. Nature neuroscience, 22(11), 1761-1770. [71]
3. Marblestone, A. H., Wayne, G., & Kording, K. P. (2016). Toward an integration of deep learning and neuroscience. Frontiers in computational neuroscience, 94. [72]
4. Richiardi, J., Achard, S., Bunke, H., & Van De Ville, D. (2013). Machine learning with brain graphs: predictive modeling approaches for functional imaging in systems neuroscience. IEEE Signal processing magazine, 30(3), 58-70. [73]
5. Vu, M. A. T., Adalı, T., Ba, D., Buzsáki, G., Carlson, D., Heller, K., ... & Dzirasa, K. (2018). A shared vision for machine learning in neuroscience. Journal of Neuroscience, 38(7), 1601-1607. [74]

8 Conclusion

Your conclusion here

9 Bonus resources

Your conclusion here

Acknowledgments

References

[1] Jon Scott Armstrong. *Principles of forecasting: a handbook for researchers and practitioners*, volume 30. Springer, 2001.

[2] Peter I Frazier. A tutorial on bayesian optimization. *arXiv preprint arXiv:1807.02811*, 2018.

[3] Martin Pelikan, David E Goldberg, Erick Cantú-Paz, et al. Boa: The bayesian optimization algorithm. In *Proceedings of the genetic and evolutionary computation conference GECCO-99*, volume 1, pages 525–532. Citeseer, 1999.

[4] Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P Adams, and Nando De Freitas. Taking the human out of the loop: A review of bayesian optimization. *Proceedings of the IEEE*, 104(1):148–175, 2015.

[5] Jasper Snoek, Hugo Larochelle, and Ryan P Adams. Practical bayesian optimization of machine learning algorithms. *Advances in neural information processing systems*, 25, 2012.

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[7] Michael I Jordan and Tom M Mitchell. Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245):255–260, 2015.

[8] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.

[9] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. PMLR, 2015.

[10] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28, 2015.

[11] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[12] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.

[13] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *nature*, 518(7540):529–533, 2015.

[14] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[15] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[16] Stuart Rusell and Peter Norvig. Artificial intelligence: A modern approach. *Pretice Hall Series in Artificial Intelligence*, 1:649–789, 2003.

[17] Tom M Mitchell. *Machine learning*. McGraw-hill New York, 1997.

[18] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

[19] Joel Grus. *Data science from scratch: first principles with python*. O'Reilly Media, 2019.

[20] Tom Hope, Yehezkel S Resheff, and Itay Lieder. *Learning tensorflow: A guide to building deep learning systems*. " O'Reilly Media, Inc.", 2017.

[21] Laura Graesser and Wah Loon Keng. *Foundations of deep reinforcement learning: theory and practice in Python*. Addison-Wesley Professional, 2019.

[22] Robin M Schmidt, Frank Schneider, and Philipp Hennig. Descending through a crowded valley-benchmarking deep learning optimizers. In *International Conference on Machine Learning*, pages 9367–9376. PMLR, 2021.

[23] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

[24] Tianqi Chen, Tong He, Michael Benesty, Vadim Khotilovich, Yuan Tang, Hyunsu Cho, Kailong Chen, et al. Xgboost: extreme gradient boosting. *R package version 0.4-2*, 1(4):1–4, 2015.

[25] Irina Rish et al. An empirical study of the naive bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence*, volume 3, pages 41–46, 2001.

[26] Gerald Forkuor, Ozias KL Hounkpatin, Gerhard Welp, and Michael Thiel. High resolution mapping of soil properties using remote sensing variables in south-western burkina faso: a comparison of machine learning and multiple linear regression models. *PloS one*, 12(1):e0170478, 2017.

[27] Dastan Maulud and Adnan M Abdulazeez. A review on linear regression comprehensive in machine learning. *Journal of Applied Science and Technology Trends*, 1(4):140–147, 2020.

[28] Evangelia Christodoulou, Jie Ma, Gary S Collins, Ewout W Steyerberg, Jan Y Verbakel, and Ben Van Calster. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *Journal of clinical epidemiology*, 110:12–22, 2019.

[29] Stephan Dreiseitl and Lucila Ohno-Machado. Logistic regression and artificial neural network classification models: a methodology review. *Journal of biomedical informatics*, 35(5-6):352–359, 2002.

[30] Thomas G Dietterich and Eun Bae Kong. Machine learning bias, statistical bias, and statistical variance of decision tree algorithms. Technical report, Citeseer, 1995.

[31] William S Noble. What is a support vector machine? *Nature biotechnology*, 24(12):1565–1567, 2006.

[32] Lipo Wang. *Support vector machines: theory and applications*, volume 177. Springer Science & Business Media, 2005.

[33] Marti A. Hearst, Susan T Dumais, Edgar Osuna, John Platt, and Bernhard Scholkopf. Support vector machines. *IEEE Intelligent Systems and their applications*, 13(4):18–28, 1998.

[34] Leif E Peterson. K-nearest neighbor. *Scholarpedia*, 4(2):1883, 2009.

[35] Min-Ling Zhang and Zhi-Hua Zhou. A k-nearest neighbor based algorithm for multi-label classification. In *2005 IEEE international conference on granular computing*, volume 2, pages 718–721. IEEE, 2005.

[36] Kristina P Sinaga and Miin-Shen Yang. Unsupervised k-means clustering algorithm. *IEEE access*, 8:80716–80727, 2020.

[37] Gérard Biau and Erwan Scornet. A random forest guided tour. *Test*, 25(2):197–227, 2016.

[38] Qianshi Wei, Ying Jiang, and Jeff ZY Chen. Machine-learning solver for modified diffusion equations. *Physical Review E*, 98(5):053304, 2018.

[39] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in Neural Information Processing Systems*, 34:8780–8794, 2021.

[40] Leslie Pack Kaelbling, Michael L Littman, and Andrew W Moore. Reinforcement learning: A survey. *Journal of artificial intelligence research*, 4:237–285, 1996.

[41] Richard S Sutton, Andrew G Barto, et al. *Introduction to reinforcement learning*. MIT press Cambridge, 1998.

[42] Csaba Szepesvári. Algorithms for reinforcement learning. *Synthesis lectures on artificial intelligence and machine learning*, 4(1):1–103, 2010.

[43] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*, 2013.

[44] Kotaro Hara, Abigail Adams, Kristy Milland, Saiph Savage, Chris Callison-Burch, and Jeffrey P Bigham. A data-driven analysis of workers' earnings on amazon mechanical turk. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pages 1–14, 2018.

[45] Alexandru Niculescu-Mizil and Rich Caruana. Predicting good probabilities with supervised learning. In *Proceedings of the 22nd international conference on Machine learning*, pages 625–632, 2005.

[46] Rich Caruana and Alexandru Niculescu-Mizil. An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd international conference on Machine learning*, pages 161–168, 2006.

[47] Thomas Hofmann. Unsupervised learning by probabilistic latent semantic analysis. *Machine learning*, 42(1):177–196, 2001.

[48] Xiaojin Jerry Zhu. *Semi-supervised learning literature survey*. University of Wisconsin-Madison Department of Computer Sciences, 2005.

[49] Olivier Chapelle, Bernhard Scholkopf, and Alexander Zien. Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews]. *IEEE Transactions on Neural Networks*, 20(3):542–542, 2009.

[50] Dong-Hyun Lee et al. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *Workshop on challenges in representation learning, ICML*, volume 3, page 896, 2013.

[51] Fjodor Van Veen and Stefan Leijnen. The neural network zoo. *The Asimov Institute*, 2016.

[52] Stefan Leijnen and Fjodor van Veen. The neural network zoo. *Multidisciplinary Digital Publishing Institute Proceedings*, 47(1):9, 2020.

[53] Anil K Jain, Jianchang Mao, and K Moidin Mohiuddin. Artificial neural networks: A tutorial. *Computer*, 29(3):31–44, 1996.

[54] Mohamad H Hassoun et al. *Fundamentals of artificial neural networks*. MIT press, 1995.

[55] Pawel Kudela, Piotr J Franaszczuk, and Gregory K Bergey. Changing excitation and inhibition in simulated neural networks: effects on induced bursting behavior. *Biological cybernetics*, 88(4):276–285, 2003.

[56] Sumit Goyal and Gyanendra Kumar Goyal. Application of simulated neural networks as non-linear modular modeling method for predicting shelf life of processed cheese. *Jurnal Intelek*, 7(2):48–54, 2012.

[57] Zachary C Lipton, John Berkowitz, and Charles Elkan. A critical review of recurrent neural networks for sequence learning. *arXiv preprint arXiv:1506.00019*, 2015.

[58] Chuanlong Yin, Yuefei Zhu, Jinlong Fei, and Xinzheng He. A deep learning approach for intrusion detection using recurrent neural networks. *Ieee Access*, 5:21954–21961, 2017.

[59] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

[60] Xin Yi, Ekta Walia, and Paul Babyn. Generative adversarial network in medical imaging: A review. *Medical image analysis*, 58:101552, 2019.

[61] Alankrita Aggarwal, Mamta Mittal, and Gopi Battineni. Generative adversarial network: An overview of theory and applications. *International Journal of Information Management Data Insights*, 1(1):100004, 2021.

[62] Saad Albawi, Tareq Abed Mohammed, and Saad Al-Zawi. Understanding of a convolutional neural network. In *2017 international conference on engineering and technology (ICET)*, pages 1–6. Ieee, 2017.

[63] Keiron O'Shea and Ryan Nash. An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*, 2015.

[64] Ruslan Salakhutdinov and Iain Murray. On the quantitative analysis of deep belief networks. In *Proceedings of the 25th international conference on Machine learning*, pages 872–879, 2008.

[65] Geoffrey E Hinton. Deep belief networks. *Scholarpedia*, 4(5):5947, 2009.

[66] Gail A. Carpenter and Stephen Grossberg. The art of adaptive pattern recognition by a self-organizing neural network. *Computer*, 21(3):77–88, 1988.

[67] Radu Dogaru and Ioana Dogaru. Lb-cnn: An open source framework for fast training of light binary convolutional neural networks using chainer and cupy. *arXiv preprint arXiv:2106.15350*, 2021.

[68] Coline Devin, Abhishek Gupta, Trevor Darrell, Pieter Abbeel, and Sergey Levine. Learning modular neural network policies for multi-task and multi-robot transfer. In *2017 IEEE international conference on robotics and automation (ICRA)*, pages 2169–2176. IEEE, 2017.

[69] Bart LM Happel and Jacob MJ Murre. Design and evolution of modular neural network architectures. *Neural networks*, 7(6-7):985–1004, 1994.

[70] Neil Savage. How ai and neuroscience drive each other forwards. *Nature*, 571(7766):S15–S15, 2019.

[71] Blake A Richards, Timothy P Lillicrap, Philippe Beaudoin, Yoshua Bengio, Rafal Bogacz, Amelia Christensen, Claudia Clopath, Rui Ponte Costa, Archy de Berker, Surya Ganguli, et al. A deep learning framework for neuroscience. *Nature neuroscience*, 22(11):1761–1770, 2019.

[72] Adam H Marblestone, Greg Wayne, and Konrad P Kording. Toward an integration of deep learning and neuroscience. *Frontiers in computational neuroscience*, page 94, 2016.

[73] Jonas Richiardi, Sophie Achard, Horst Bunke, and Dimitri Van De Ville. Machine learning with brain graphs: predictive modeling approaches for functional imaging in systems neuroscience. *IEEE Signal processing magazine*, 30(3):58–70, 2013.

[74] Mai-Anh T Vu, Tülay Adalı, Demba Ba, György Buzsáki, David Carlson, Katherine Heller, Conor Liston, Cynthia Rudin, Vikaas S Sohal, Alik S Widge, et al. A shared vision for machine learning in neuroscience. *Journal of Neuroscience*, 38(7):1601–1607, 2018.