Q/A @ CCN

January 27, 2016

Below all the questions are listed, grouped approximately by topics and some of them have short answers inspired by the session. The answers are far from complete etc.. For more info, contact vadim.kulikov@helsinki.fi.

1 ANN's

1. It appears there are a lot different models for artificial neural networks. Can they be classified somehow (even on a broad scale), and if they can, into which categories?

   A: Usually according to architecture: feedforward, recurrent, continuous time, discrete time. And according to learning: supervised, unsupervised, reinforcement. Also according to primitive functions: smooth sigmoid, half-linear ($f(x) = 0$ for $x < 0$ and $f(x) = x$ for $x \geq 0$) etc..

2. A follow-up to the previous question - it would seem logical that different models are used for different tasks. Can you name a few of these and corresponding applications?

   A: Feedforward networks are mainly used for classification (e.g. images). Tasks requiring temporal dynamics and internal memory are better dealt with recurrent NN's, for example natural language processing. Continuous time NN's have more "expressive power" in comparison to discrete time NN 's since the network has more possible states. CTRNN are applied widely from natural language processing to cognitive modelling.

   http://arxiv.org/pdf/1305.1958.pdf

3. What is the difference between Neural Networks and Multilayer perceptrons?

   A: A Multilayer perceptron is just one type of a neural network which is trained in a particular way.

4. How do we choose which functions to use in an artificial neural network? What are the main features of the used functions to be taken into account while designing an artificial neural network?

A: I assume you mean the integration of the input signals and producing output. Most commonly the integration is linear with a subsequent non-linearity applied. What is this non-linearity is often determined by the needs of the algorithms, its efficiency, and simply experience of the programmer. For example applying gradient descent learning requires a differentiable function such as a smooth sigmoid. On some other cases a step function is more appropriate, because it is easier to handle. In some cases the half-linear $f(x) = 0$ for $x < 0$ and $f(x) = x$ for $x \geq 0$ provides better convergence.

2 ANN's vs other algorithms/machines

1. What are the most prominent strengths and weaknesses of neural networks, as compared to other machine learning techniques (for example logistic regression, random forests, support vector machines).

   A: Neural networks are typically more general than any of those. For example SVM can (usually) be modelled with one neuron or one-layer network.

   Logistic regression is very good for solving simple classification problems, but is not good for dealing e.g. with complex visual scene-recognition. There are many different architectures of NN's which are flexible and "stackable": one neural network can be attaced to another to make a new one etc.

   "NN are the second best thing for everything"

   Random forest is the first thing to try in a classification problem, but in certain things NN's are overwhelmingly better (such as image classification).

2. Why have neural networks apparently enjoyed more success than other biologically inspired algorithms, such as genetic programming?

   A: Neural networks are often combined with other learning algorithms such as evolutionary learning.

   In 1960's people thought that perceptron is a breakthrough (NY Times: "An embryo for a conscious robot"), but it was an optimistic illusion.

   Genetic algorithms are also popular, but maybe with a smaller range of applications.

   NN's converge faster; algrorithms are precise in detecting the direction of convergence/error compared to e.g. genetic programs.

3. How neural networks differ from other computing models (lambda calculus, turing machines, cellular automata)?

   A: Essentially a NN is a cellular automaton, unless it is continuous time..

4. What set of primitive functions is required to make artificial neural networks as powerful as a Turing machine?

   A: See the answers to the next question.

5. Are neural networks Turing-complete? In more practical terms: are there problems neural networks cannot solve? Are there ones that can only be solved by neural networks but not "regural" by algorithms?

   A: A discrete time recurrent NN with rational weights and sigmoid activation function is Turing complete in the following sense: for each recursive function f there is a finite NN which with input "n" outputs "$f(n)$"

http://www.math.rutgers.edu/~sontag/FTP_DIR/aml-turing.pdf

A: If real values are allowed, super-Turing power is obtained: Balcazar, J.L.: "Computational power of neural networks: a characterization in terms of Kolmogorov complexity".

A: One layer continuous NN can approximate any continuous function on a compact set. (This is similar to Taylor/Fourier series).

6. "Although neural networks and cellular automata are potentially more efficient than conventional computers in certain application areas, at the time of their conception they were not yet ready to take center stage. The necessary theory for harnessing the dynamics of complex parallel systems is still being developed right before our eyes. In the meantime, conventional computer technology has made great strides" What is the state of the art of this research? Who are the leading scientists?

   A: Perhaps you should look at research on parallel computing

   http://goparallel.sourceforge.net/

7. Rojas chapter 1 mentions that "digital signals combined in an excitatory or inhibitory way can be used to implement any desired logical function". How would one go about implementing a NAND -gate as a neural network?

   A: 3 neurons: 2 input one output. Weight matrix=[-2,-2], threshold of output = -3. Alternatively add one neuron with constant output 4 to avoid negative threshold.

3 Modelling

1. Traditionally when discussing the brain in context of computation we are only interested in neurons. Years ago I read from some pop-science mag that glial cells are actually much more important than we originally thought of, and not just some sort of support structure. So the question is: are we losing something when we consider only neurons or is there a good reason why we can discard the glial cells (and whatever else there is in the brain in addition to neurons)?

   A: The function of gliacells is not to my knowledge adopted to computational neuroscience. Also, very little is known about it.

2. If I understood correctly, the usual computational neural network models of neural system model just the 'network' of neural cells. What about all rest of "messy" biological and chemical side of things, e.g. certain drugs that function by affecting brain chemistry? Are there models for these phenomena, too?

3. What are the most common simplifications used in artifical neural networks compared to biological neural networks and what limitations or side effects do they produce?

   A (covering the 1-3 above): There are two ways of looking at this: normative and descriptive. Normative: what is the best way to implement this and that? In this case we are only inspired by the brain, but are not aiming to model it precisely. Then a lot of simplifiations are made without worry. Descriptive: trying to make realistic models. This is very hard. In the Human Brain Project, the simulation of 1 neuron takes the full power of a modern laptop and still many neuroscientists think it is way too simplistic. Models of hormonal and pharmaceutical action exist (see neuropharmacology), but are themselves complicated models of specific functions of the brain such as modulation of theta-wave activity in hippocampus.

   http://www.scholarpedia.org/article/User:P%C3%A9ter_%C3%89rdi/Dynamic_neuropharmacology

4. Is information stored as binary (or discrete values) or as real (or continuous values) in nerve cells?

   A: Neither. The brain stores information in physical properties of nervous cells and their connections. Most of these things are best modelled using real number continuum, so in that sense the answer "real values" would be closer to correct. Each synapse can be modified in multitude of different ways specified by the biochemistry of it. However, most of the important signals (action potentials) are all-or-none-phenomenon. But here again, the *frequency* of firing is a continuous phenomenon.

5. I'm interested about relatively small neural networks, for example something like insects neural system, they appear to be easier to study than human brain, nevertheless behavior of insects can be quite complex. r there any noteworthy major breakthroughs in this kind of a research, just generally speaking?

   A: Minimal networks in cognitive modelling modelling:

   Beer et al:

   http://onlinelibrary.wiley.com/doi/10.1111/cogs.12142/full,

   Froese et al:

   http://arxiv.org/pdf/1305.1958.pdf

   Smallest nervous system mapped: C. Elegance.

   http://www.openworm.org/

6. In Rojas book there's analogy between neurons communication and frequency modulation, I know what frequency modulation is but I didn't get the analogy, may be discussion will help me to get the idea better

   A: Neuron outputs a continuous number (a real number) in the form of frequency of action potentials. Suppose a neuron wants to output number π. Then it fires π spikes per second...

7. Has there been any attempts to replicate specific animal brain neural topologies with artificial neural networks, for example using an mri etc to figure out the topology of a cat brain, and then replicate the same topology with artificial neurons. (I see a problem with figuring out the weights etc for the connections, but you never know)

   A: We are very far from cat, but not so far from the smallest nervous system in the world: C. Elegance

   http://www.openworm.org/

8. In this course, we concentrate on sight (visual system). What about sound / hearing? To which extent same models are applicable? Any pointers to material regarding that?

   A: There is little I can say about this. Many methods probably apply, like NN's. But sound seems to be quite a bit more complicated than vision. For example melody recognition is where visual recognition was 30 years ago.

9. In your opinion, how does brain do unsupervised learning ? More precisely, do you really think classic methods like ICA, PCA, Denoising Source Seperation (DSS), Self Organizing Map or Ladder Networks implemented explicitly in our brains ? If so, how are they implemented.

   A:

http://www.scholarpedia.org/article/User:P%C3%A9ter_%C3%89rdi/Dynamic_neuropharmacology#Neurobiological_and_psychiatric_disorders_as_dynamical_diseases

http://www.sciencedirect.com/science/article/pii/S0893608099000465

4 Ion channels

1. Why potassium has a negative equilibrium potential while sodium has a positive one, given that they both carry positive charges? Is it because most potassium ions are initially inside the cell but most sodium ions are outside? The corresponding material is on the 13 page of Roja's book.

   A: Difference between resting potential and equilibrium potential. THe equilibrium potential depends on the consentrations outside and inside of the cell. Sodium is kept mostly outside and is not in an equilibrium while potassium is almost in equilibrium in the resting state.

2. Why do the ion channels operate the way they do (sodium channels opening at threshold potential, potassium channels opening later)?

3. I don't understand why potential difference is only related to the ion concentration based on the Nernst formula. It seems to me it should be related to the number of positive and negative charges. The corresponding material is on the 13 page of Roja's book.

   A: It has nothing to do with other charges than the ion itself, but it is related to the number of ions, k depends on that.

4. Exercise 3.1: cell membrane potential. I see that there has to be a suitable mixture of both ions to achieve that potential, but I can't explain the mechanism why this happens. How does it work?

   A: We consider exercises on Thursday........... ion pumps.

5 Axons and Synapses

1. If the axon of a neuron connects to more than one target neuron, is the firing rate the same for each target?

   A: yes. The action potential splits at the stems and continues in the same way in both.

2. Does the amplitude of the emitted impulses of a neuron vary and does it matter?

   A: Action potential has always the same amplitude: from -70 to +40 mV.

3. In Rojas, it was stated that information is stored in synapses and that there are other ways of storing information. However, these other ways were not discussed. What other ways of storing information are there in the brain?

   A: 1. geometry of neurons 2. short term information may be stored in the temporal differences in the oscillation of neurons, 3.

4. How do synapses determine the direction for the transmission of information? That is, why signals flow from one cell to another in a well-defined manner and not both-ways? How this direction is determined?

   A: The hyperpolarisation after the action potential, which is the result of potassium coming out of the cell, is the reason for a refractory period. Also the channels don't work for a while after the action potential. That's why the signal cannot change direction.

5. How are synapses account for in the theory of (artificial) neural networks?

   A: Weights of the connections.

6. If a neuron is on a refractory period, where it does not activate due to input from other neurons, does it still fire at the base rate or is its activity actually decreased from its base rate?

   A: the refractory period is shorter than the time between two action potentials in the base rate, so not really applicable.

7. what does the Hodgkin-Huxley DE describe?

   A: How exactly do the voltage gated channels activate and when during the action potential.

8. Why are there more chemical than electrical synapses?

6 Neurons and Brain

1. Explain or discuss information transmission of neurons.

2. Can neuroscience gain something out of the fact that there are different Action potential depending on the cell (like heartaction potential is different) or is there any reason why only the "normal" Action potential is explained?

   A: Cognitive neuroscience is mainly interested in the way information is processed. Neural action potentials deliver information from one cell to another. If some other processes transmit or store or process information such as cardiac action potentials etc, then they should also be of interest. The only type of "other" action potential I have seen considered in neuroscientific context is the dendritic action potential which is apparently of high importance:

   https://en.wikipedia.org/wiki/Dendritic_spike

3. How does Microglia repair neurons? And what kind of potential has that mechanism in Neuroscience?

   A: Microglia seem to be responsble for cleaning the brain tissue of dead neurons and other uncleanliness. The details and other functions are beyond my current knowledge. Macroglial cells produce myelin which an insulator around axons and makes it possible for the action potential to travel faster. They are also more numerous than microglia and are probably more important. Recently it has been suggested that glia might take part in information processing of the brain: namely they affect breathing. This is logical, because if their function is to "take care" of neurons, they should also make sure that they receive enough oxygen. Some gliacells were demonstrated to influence breathing signals

   http://discovermagazine.com/2011/jan-feb/62

   This kind of thing has not, however, been yet adopted to computational cognitive neuroscience.

4. On Wikiversity site, different divisions of the brain were discussed. Are there large differences in the level of neurons in between these brain parts?

   A: The trend seems to be that the neuronal density is higher in the back of the head and lower towards the front. With some exceptions:

   http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2936588/

   http://www.pnas.org/content/113/3/740.full

   However, the "complexity of pyramidal cells increases from posterior to anterior brain regions."

   (https://en.wikipedia.org/wiki/Pyramidal_cell)

5. What is the current level of research (any word on a particular subfield is interesting) on "computational primitives" mentioned in Rojas' first chapter? Has there been recent research outside of neural networks for example? This might be a broad question, as I'm not sure what there is at the moment, but the topic is interesting.

6. This one's a bit more informal, but I noticed the NIS-material is lacking in explaining linear algebra (chapter 19). I wonder if someone would need another material for support, so if there is a coherent place to look for you could recommend?

   A: This material is very clear and has the right content:

   http://genes.mit.edu/burgelab/yarden/strang_linear_algebra.pdf

7. How specialised/general are the brain areas

8. What's the "frequency" of the brain? Starting from a "typical" firing frequency for a neuron in inactive mode, if we knew the depth of the neural network (or some approximation of it) we could estimate the time required for one cycle and get some sort of estimate of the capacity of the brain.

   A: I am not sure what this means and how we get to capacity... but: EEG measures the so called brain waves. Neurons are constantly oscillating more or less in synchrony. The frequences are from 1 Hz (delta oscillations in deep sleep) to 40 Hz (gamma waves).

9. Biological neurons have a base-firing rate, where they continuously send action potentials at certain intervals regardless of input. Why do they do that? Does it serve any known purpose?

10. How many nerve cells can a nerve cell be linked at once?

    A: thousands to tens of thousands, on average 7000 synapses per neuron in human brain. This means that one neuron has on average 7000 incoming and 7000 outgoing connections, in total connected to 14000.

11. If the synapses function in an "all-or-none" fashion, how does the summation of action potentials at the axon hillock work?

    A: The amount of excitation or inhibition provided by incoming actino potentials is not on/off, but a continuous phenomenon which depends on the position of synapse on the dendrite, the post-synaptic receptors, the neurotransmitters, on the temporal relationship of incoming signals etc etc. When the excitation crosses a certain threshold, the cell fires an actino potential.

12. How does long term memory work at neuron level?

    A: LTP, post-synaptic density, dendrite geometry..

13. Being able to store information (beyond simple input to dendrites), does it affect how processing is done at neuron level?

7 Statistics and Math

1. In which scenarios would you use maximum likelihood and when maximum a posteriori?

   A: if the prior is 'flat', then they are connected by a constant.

2. What is a good method to compute the maximization of the likelihood/posterior numerically, if there is no analytical solution?

   A: Gradient ascent, chapter 18 from NIS.

3. Since Fourier transforms / series are apparently important in visual processing, could we get a short recapitulation of the math behind that?

   A: In the beginning of the second part of the course.

4. What is the expectation of the variance $E(z1 - Ez1)^2$? Is it the mean value of the variances of the random variable z1?

   A: $E((z1 - Ez1)^2)$ is by definition the variance. Variance is a number, not a random variable.

5. What is a primitive recursive function? Why are we interested in those?

   A: How is this related to this course? Go to 'Matemaattinen logiikka' upstairs.

6. What is Bayesian inference? What can you calculate with it?

   A: We will cover some basics in the last week of the course. Mainly, Bayesian inference gives a statistically optimal way to update an internal belief about the world (which is represented in terms of a probability distribution) using prior belief and new evidence as starting point. Cool applications:

   http://www.sciencedirect.com/science/article/pii/S0960982213011287

   http://www.ncbi.nlm.nih.gov/pubmed/11807554

7. Consider formula 1.1 from NIS : It says "Generative models use Eq. (1.1) as a starting point. They attempt to explain observed data by some underlying hidden (latent) causes or factors about which we have only indirect information." How does the Equation distinguish between latent variable and noise present in the input image.

8 Other

1. Where could I read about the basics of Python? The tutorial referenced in the course home page is good, if you already know something about Python syntax etc., but I don't. Could we have a 20 minute lecture of "Python for Java programmers"? How are languages alike, which are basic differences of notation etc.?

2. What is it that makes Python such a popular language for ML purposes? (Obviously it's partly due to scipy and numpy, but are there e.g. properties inherent to the language itself that make it particularly suitable?)

3. Why do dreams only happen during REM sleep?

   A: There is some evidence that we might dream also in deep sleep, but the dreams are less conceptual and we don't remember them. According to some scientists, evolutionarily the REM-state is a much more ancient state of mind, so to speak, than our regular awakeness; namely more ancient than our cerebral cortex. The main message of this paper is that the mammalian wakefulness (and NREM) are novel, cortical, states, largely missing in poikhilothermic animals, like reptiles, fishes, etc. REM is something much more ancient mostly requiring brainstem, and subcotical structures:

   http://www.ncbi.nlm.nih.gov/pubmed/25859205

4. What are some hot topics/new discoveries in the field that might be too recent to be found in the material, but it would be useful to be aware of?

5. I have problems understanding figure 1.17 (Rojas), and the explanatory text associated to it. Especially, what are the phase factors d0..dn?

   A: The neuron collects numbers that it receives and adds them together. Then it applies the sin function to the result. Suppose N_0 is a neuron receiving information x from only one connection with weight a and with phase factor d. Then it outputs $\sin(ax+d)$. If $d = 0$, this equals $\sin(ax)$, if $d = \pi/2$, then $\sin(ax + d) = \cos(ax)$. So by regulating d, one can influence whether the neuron corresponds to sin or cos in the Fourier series.

6. What exactly is a feature detector? I have been trying to find a formal definition, but wasnt able to find one.

   A: It's like Jazz, you can't define it, but you know when you see one.