DD2437 Online Re-exam Questions, 21HT (2021/12/17)

Part I - Quiz (23p)

Q1 (2p)

Please mark correct statements:

a) the problem of dead units is typically mitigated by greedy learning in competitive networks

b) formula for the network output in RBFs and MLPs is the same, i.e. $y = f(\Sigma_i w_i x_i)$

c) in Hopfield networks, performing sequential update speeds up the convergence as opposed to using batch update

d) discriminative neural networks tend to be faster to train than generative models to reach a comparable level of classification performance

e) storage capacity in associative memory networks depends on the learning rule

f) committee of networks have better generalization potential than individual networks due to the emerging reduction of bias in ensemble learning

g) the main computational advantages of long short-term memory networks (LSTM) over vanilla recurrent neural networks (RNNs) lies in the depth of the underlying architecture

h) backprop can be used for unsupervised learning of data representations

Q2 (1p)

In which of the following cases it is common to use the *least means squares rule* for learning the weights?

a) for the weights between the input layer and the RBF-layer in an RBF-network

b) for the weights between the input nodes and the hidden nodes in a Restricted Boltzmann Machine (RBM)

c) for the weights of the recurrent connections between nodes in a Hopfield network

d) for the weights between the RBF-layer and the output layer in an RBF-network

e) for the weights between the input nodes and the reservoir nodes in an Echo state network

f) for the weights of the fully connected hidden layers near the outputs in convolutional neural networks (CNNs)

Q3 (3p) *(19vt)*

a) For a single-layer perceptron, what are the convergence criteria of the perceptron learning rule and delta rule? List key differences between the two learning algorithms and discuss briefly the major weakness of the original perceptron learning.

b) Then, perform computations on the following initial weight matrix **W** (the last column contains the bias; the threshold for the step activation function is 0)

$$\mathbf{W} = \begin{bmatrix} 1 & -3 & -1 & -1 \end{bmatrix}$$

subject to one epoch of batch training using the delta rule with two patterns, p1 and p2, belonging to the "positive" class (+1) and two patterns, p3 and p4, representing the "negative" class (-1):

$p1 = [\,2, 0, -1]^T$, $p2 = [1, 1, 2]^T$;

$p3 = [\,4, 1, -1]^T$, $p4 = [\,0, 1, -1]^T$.

In your b) answer please provide **the final weight matrix** after the requested training with the learning ret of 0.1 and indicate **which patterns are correctly classified**.

You should type both answers, a) and b), in the box.

Q4 (3p) (*20ht*)

Briefly address each of the three following instructions, I-III:

I) Please explain the intuition behind the use of weight decay as a regularisation technique (you do not have to explain how it is implemented, focus on why preventing long learning is thought to be desirable)

II) Why do we consider generalisation to be a fundamentally important capability of neural networks? What are key factors that determine the potential of a neural network to generalise?

III) List three distinct examples of techniques that help neural network designers to reduce the variance in the bias-variance dilemma. Explain for each of them the underlying concept in one sentence.

You should type in your answer in the box.

Q5 (2p) *(20vt)*

The process of learning neural networks is generally considered as stochastic, which may be interpreted in multiple ways depending on concrete problems and circumstances. Please indicate whether the statements below capture right interpretation in given contexts and are thus correct (mark TRUE statements):

a) The only factor introducing randomness to the process of tuning multi-layer perceptron (MLP) weights with online backprop is random weight initialisation (given a fixed model and other hyperparameters).

b) Deep belief networks (DBNs) are considered as stochastic networks due to a probabilistic description of weights.

c) Ensemble learning helps minimising different sources of variability attributed to weak learners.

d) Generative adversarial networks (GANs) provide a model of data probability distribution by estimating the parameters of the distribution (often Gaussian).

e) Cross-validation as a random sampling method allows for estimating the distribution of a generalisation error (not only the mean estimate).

f) Bayesian regularisation builds on explicit sampling from the probability distributions describing response properties of individual neurons.

Q6 (2p)

For each of the presented problems please indicate the single best suited neural network approach:

a) three-class classification of colour images

-> multi-layer perceptron (MLP) with three linear outputs

-> restricted Boltzmann machine (RBM) with binary units

-> Kohonen (self-organising map) map with three-dimensional (output) grid

-> convolutional neural network (CNN) with sigmoidal outputs

b) leveraging a large dataset with only 10% annotated data for a binary classification problem

-> multi-layer perceptron with backprop training

-> overcomplete autoencoder with rectifier linear output units (ReLUs)

-> deep belief network (DBN) with greedy layer-wise pretraining

-> ensemble learning with perceptrons as weak learners

c) assigning 10-dimensional data samples into two groups

-> Hopfield network with 10 nodes

-> self-organised map (SOM) with 10-dimensional input and two-dimensional (output) grid

-> perceptron with a 10-dimensional input and two output nodes

-> radial basis function (RBF) network with 10 inputs and one sigmoidal output node

d) predicting weather parameters such as temperature and atmospheric pressure every 6^{th} hour for the next 24 hours (6h, 12h, 18h, 24h ahead)

-> a multi-layer perceptron with a set of four output nodes with linear activation functions for each weather parameter to predict (so, eight output nodes collectively for temperature and pressure)

-> long short-term memory (LSTM) network with two univariate outputs (corresponding to temperature and pressure, respectively)

-> radial basis function (RBF) network with one or two hidden nodes for every one of four time points (6h, 12h, 18h, 24h)

-> deep belief network (DBN) with two sigmoidal outputs corresponding to temperature and pressure, respectively.

Q7 (2p)

Please identify all the scenarios below that raise your concern as a neural network practitioner and match them with the underlying reason for your concern (otherwise choose the option "no major concern").

a) training data for a three-class MLP classifier with two hidden layers consists of 100 samples (25 corresponding to class I, 35 – class II and 40 – class III) with dimensionality of 30 features each

- The data is highly unbalanced with respect to class labels
- No major concern
- There are too few samples for training such a large network
- Two hidden layers are not sufficient for problems with such high-dimensional input

b) training performance for a DNN is much better than test performance

- The training algorithm has not yet converged or the learning rate is too low
- Overfitting is very likely
- No major concern
- The network has excessive bias at the cost of variance

c) there are 500 samples, each one described by 50 multimodal (binary, integer, large and small real numbers) features to be used for training Kohonen map

- The test data is not representative for the problem (defined by the training data)
- Overfitting is very likely
- No major concern
- The network has excessive bias at the cost of variance

d) Perceptron does not converge and the decision boundary keeps oscillating rather than gradually converging

- There is a problem with initialisation
- Learning rate is too high
- No major concern
- The problem is not linearly separable

Q8 (3p)

Which of the following heuristics or practical tips you would strongly recommend in given scenarios? Please prioritise and choose top three recommendations.

a) to prevent from catastrophic forgetting in a Hopfield network reduce the learning rate

b) to boost your training set of limited size perform some form of data enhancement (adding noise or making some rotational or translational transforms for images to increase the amount of the training data)

c) if data samples (input patterns) are described by small numbers, significantly smaller than 1, scale them up by a scalar when using a perceptron for their classification

d) for the problem with a limited number of available data samples try to reduce their dimensionality (with one of dimensionality reduction techniques)

e) remove all attributes (features) that happen to have missing values for some data samples

f) compare competing models based on the cross-validation estimate of their generalisation error

Q9 (2p)

Please mark correct statements motivating the choice of the hyperparameter values

a) the number of units in the hidden layer of a radial basis function (RBF) network can be decided a priori based on the insights about the number of distinct data clusters

b) the number of nodes in a Hopfield network can be chosen using a grid search with the lowest validation error criterion

c) the regularisation constant controlling the weight of the penalty term in the loss function can be tuned using gradient descent during training

d) the learning rate can be partly determined based on the number of available training data

e) grid search for an optimal configuration of hyperparameters implies that hyperparameters are optimized independently in a sequence, one at a time while the other ones remain fixed.

Q10 (3p)

Please choose three statements that best characterise the functionality of specific deep neural network (DNN) architectures (or of their components)

a) deep belief networks can be greedily pre-trained using local Hebbian learning rule.

b) convolutional neural networks (CNNs) facilitate translational invariance due to weight sharing

c) gates in long short-term memory (LSTM) cells help LSTM networks to handle periodicities at different, often unknown, temporal scales

d) variational autoencoders represent a class of generative models that learn non-parametric data distribution (implicitly, without the probabilistic underlying model)

e) generator and discriminator networks that comprise a generative adversarial network (GAN) must be first pre-trained independently before they can be coupled by the joint cost function

f) end-to-end weight fine tuning of stacked autoencoders with backprop (after pre-training) requires labelled data.

Part II (23p)

Question II.1 (3p)

Music information retrieval is a dynamically growing field with many real-world applications. It has recently attracted a lot of attention in deep learning community. Yet, most applications have been centred on music classification with regard to, for example, genre. However, it is more challenging to study music more exploratively with neural network tools. Here you are requested to propose a neural network based solution to the problem of categorising music created in the so called Classical Period. You have access to digital recordings of more than 1200 pieces composed by 45 composers. In the first place you are asked to identify salient categories for the musical pieces based solely on the recordings – so far experts have identified 10-15 music styles within the Classical Period but there is little consensus between them so it becomes increasingly relevant to support them with more objective purely data-driven assessment.

Please design a neural network framework for performing such an explorative analysis. Consider the challenge of suitable data representation of the recordings and their actual grouping (proposing labels for newly formed categories). Please motivate your choices, name the algorithms involved (without any technicalities), identify key parameters that determine the "granularity" of grouping and briefly explain how your system could be used by music experts to further explore those music categories (or subcategories). In particular, the experts would like to study if there is a great deal of overlap between the categories found by your algorithm and those they proposed through their own research. Also, describe how they could quickly examine whether individual composers tend to produce music that represents more than a single category.

Question II.2 (4p)

Please determine what type of problem is posed in p. a) and b). In both cases, propose a neural network approach to address it. In particular, describe the network archiecture, characterise inputs and outputs (what kind of data, how many nodes etc), explain how training is done, etc.

a) As machines get more and more complex, the importance of self-diagnostic tools also grows. In an automated factory, a number of measurable variables (26 in total, including sensor values, data from micro-controller memories, etc) are constantly logged. Over time the company has saved this data (now about 122000 entries) including annotations labelling the data. These labels are either simple collective descriptors such as "normal operation" or, in cases of abnormal/dysfunctional machine state, the type of error was logged. Your task is to use this data to construct an automated neural network based system for diagnostics.

b) In a gaming company you are responsible for one of the virtual characters featuring in a certain game about medieval times in France. These virtual characters need to be controlled by some intelligent algorithm (gaming AI). Your character is a greedy baronet who constantly tries to accumulate wealth. Essentially, you should construct the algorithm that gives the character the capacity to maximize income. The baronet has an inspector to send out to the different estates to collect earnings, and the action of this inspector needs to be given by your system. These visits of the inspector also serve as opportunities to enforce the law and punish disobedient workers, and therefore should be as frequent as possible, and importantly no estate should be missed getting a visit. But, travel takes time and costs money, and the baronet needs to be clever in how to send out the inspector to maximize profit. Your task is to propose a neural network algorithm that helps the baronet in this effort.

Question II.3 (3p)

To study cognitive phenomena in the interdisciplinary consortium of researchers you are asked to support computational efforts aimed at modelling an abstract memory system. The requirements are that it should be capable of robust learning of new patterns and storing the existing ones. Which neural network architecture could you propose as a prototype for such a memory system and what would be the recommended learning approach? How would you retrieve/recall specific memories stored in the network? How would your network behave when new memories are to be acquired in the situation where the network capacity has approached the upper limit? What parameters and features determine the number of unique memory patterns that can be robustly stored in your model?

Question II.4 (5p)

In the elderly care there is an urgent need for autonomous systems that can monitor senior citizens and send an alarm in the case the person has fallen, appears confused or otherwise behaves different from normal. In a study, a set of sensors (ultrasound movement detectors, microphones, video cameras, etc) were installed in the home of a set of elderly. The video cameras were only used for the development of the system and in the final commercial system no video cameras were included. During the development process, the video was monitored by human experts and different situations were labelled as ”fall” (typically when the person was lying on the floor without getting up easily), ”confusion” (typically associated with erratic movement like purposeless walking from room to room) or ”normal”. The time markers corresponding to these labelled scenarios allowed for dividing the continuous sensor recordings into 8 time segments. This means that each video-segment labelled by the experts (into one of the three categories mentioned above) was described by 8 samples for each recorded sensor. In total, for all test persons/subjects 8913 time-segments were produced.

Your task is to design a neural network based identification of the person's state so that an alarm would be triggered every time "fall" is detected, or other form of support could be offered when an elderly is found to be "confused". What type of problem does the company want to address? Accordingly, propose a neural network solution in your design assignment. In particular, motivate the choice of your network type, briefly characterise its topology clearly indicating the inputs and outputs, and describe how the network should be trained – how the data should be used and what learning algorithm you recommend. Please, explain also how you would optimise and estimate the generalization capacity. Finally, identify key challenges and potential difficulties/risks concerning the problem and your approach to effectively solving it.

Question II.5 (5p)

Please design an empirical study to resolve a question as to which neural network approaches are suitable for forecasting the consumption of electrical energy in a specific residential area. Please bear in mind that the focus is on short-term prediction, max 24 hours ahead. The dataset for your study contains samples of measured energy load on an hourly basis from the last three years with large seasonality trends removed (data normalised across four seasons of the year) and remaining strong weekly as well as daily periodicity. When describing your experimental study please keep the following questions/issues in mind:

− What could be candidate neural network types for your comparative study? What configurations would you consider worth testing (how would you go about parameter settings and what parameters)? What would you consider a fair comparison in terms of network configurations?

− How would you suggest using your data, how would you use/divide these roughly 26280 (3x365x24) samples to ensure reliable training, validation and testing?

− What would you feed to the networks, what would be inputs and outputs? You could support your short description with suitable illustrations.

− What would be the criterion/criteria for your comparison (what would you measure)? How would you ensure that your findings account for stochastic nature of training neural networks so that generalizable conclusion could be drawn (rather than being an unreliable product of random effects)? How would you demonstrate your study outcomes?

Question II.6 (3p)

Please motivate briefly why we often refer to deep belief networks (DBNs) as hybrid generative and discriminative models. Explain what properties determine their generative and discriminative capabilities. Describe the process of pre-training a DBN for a simple binary discrimination task.