Old exam, Part 2 are answered

Q1:

This seems to be an unsupervised grouping task where we want to explore the data and maintain topological ordering. For this, a SOM network trained in competitive learning however in a sense where the neighbours are allowed to learn as well (leaky learning), 2-dimensional grid. The neighbourhood size is to be found by model selection and its decaying factor as well. Learning rate is also a hyperparameter that matters. The amount of clusters that pop up represents the total genres during that period. The data representation could be a frequency map, or time-series. To check if a composer falls into different genres, simply input some songs of that composer and check if the song falls into nearby categories or far away,

Q2

1. This seem to be two types of problems. Either anomaly detection or classification task. The model suggested would be a feed-forward MLP. Either train it on all the sensors (multimodal) or train it on all single 26 sensors to classify. The learning rate and hidden nodes in each hidden layer and hidden layer is up for model selection through cross-validation of the entries. Training through back-prop minizing cross entropy loss, output layer would be depending on the classes and done with sigmoid. Normalize each data input channel.

2. Seems to be an optimization problem similar to that of TSP problem. Thereby Hopfield, or SOM, network can be used but in case use Hopfield. Bipolar encoding where unvisited are -1 and visited are 1, punish repeated visits, longer distances and hard cost in missing states iin energy minimization. Let the Hopfield network energy minimise this and stabilize on a solution.

Q3

Network: Hopfield networks
Learning: Hebbian Learning
Recall: Conducted by providing the input corresponding to the memory pattern itself or a noisy/incomplete version and performing iterative updates until the network converges. The network is auto-associative. Networks trainde with hebbian learning suffers from catastrophic forgetting, if upper limit is breached then the model will forget almost everything. Parameters that determine the number of unique memory patterns are nodes in the network and how similar the data points are to each other.

Q4

What kind of problem: Seems to be a classification of 5 classes.
Network: Seems to be an MLP feed-forward network. The amount of hidden layers and hidden nodes is to be decided by the model selection.
Input: The data representations of the eight sensors
Output: Sigmoid function of which class the patient best falls into
Learnig: Backpropp algorithm however with some regularisation as there seems to be a lot of data and to prevent overfitting and improve generalization.
Others: User cross validation method to check generalization error.
Challenge: Unbalanced dataset, class " Normal" will be here a lot.

Q5 - LSTM - Questi

- **What neural network:**

MLP with time-lagged representations.

- **How to divide the dataset:**

- **What would you feed into the networks?**

- **What would be the criterion for your comparison**

Q6 - DBN - restudy what makes them special

Why do we often refer to DBN as hybrid generative and discriminate models?

- DGN models can be used as for generative and discriminative tasks at the same time with the same model. You do not need to modify the model

What properties determine their generative and discriminative capabilities

- Generative capabilities are owed to a probabilistic representation of the joint (input + output= data distribution with the support of latent variables.

- Discrimination is facilitated by modelling the projections to the class label layer and thus estimating class conditional probabilities P(Output| Input)

Process of pre-training a DBN for a simple binary discrimination task

- DBN consists of multi-layers of Restricted Bolzman machines. You train the first RBM layer which takes in the input data. Trained with Contrastive Divergence with gibbs sampling. The training minimizes the reconstruction error between the input data and the data generated by the RBM

- The output of the previous layer becomes the input for the next layer and this goes in. This goes on till the last layer

- Afterwards fine-tuning with supervised learning is conducted. Backpropagation or other optimization algorithms are used to minimise the classification error on the training dataset. The loss function is cross-entropy loss

- Evaluation is done with F1 score, precision, recall and other techniques.