Instructions

- The exam contains 14 questions, some of which have subquestions, for the total of 65 points.
- Do not forget to sign your name at the top of this page. **You are not allowed to take the exam questions with you.** You must return the signed exam question sheets together with the answer sheets.
- The number of points is indicated next to every question or subquestion.
- **Keep your answers as short as possible and to the point**.
- Writing long answers to demonstrate knowledge that is outside the scope of the question will not be of help. **It will cost you valuable time and may result in negative points**.

Exam: 8DM50 Machine learning in medical imaging and biology

1. You are given a dataset of features $\mathbf{X}$ and continuous target variables $\mathbf{y}$. Each sample is represented with M features and there are in total N samples the in the dataset.
   (a) (1 point) What is the dimensionality of $\mathbf{X}$ and $\mathbf{y}$?
   (b) (4 points) Outline the general idea for finding the least-squares solution of a linear regression model given a dataset $\mathbf{X}$ and $\mathbf{y}$.
2. The k-Nearest Neighbours (k-NN) classifier.
   (a) (1 point) The effective number of parameters of the k-NN classifier is:
      A. Equal to k
      B. Proportional to k
      C. Inversely proportional to k
   (b) (2 points) Motivate your answer to the previous subquestion.
3. (2 points) You are given a dataset where each patient is represented with a number of biomarkers and the target variable is a diagnosis. Is k-NN or a linear classification model easier to interpret for clinicians? Motivate your answer.
4. (5 points) What is the typical relationship between the capacity and the generalisation error of a machine learning model? Make a sketch and motivate your answer. Describe how the generalisation error when measured as the mean squared error (MSE) can be decomposed. Which component increases with increasing capacity and which component decreases?

5. Model evaluation.
   (a) (2 points) Describe in brief how the receiver operator characteristic (ROC) curve for a given binary classifier is computed.
   (b) (3 points) You want to compare the performance of your computer-aided diagnosis (CAD) method for diagnosing breast cancer from mammography images to an expert radiologist and to another competing CAD method. The two CAD methods are based on binary classifiers. The radiologist only provides hard "cancer" vs. "no cancer" diagnosis. How can you do this comparison using ROC curves?

6. (5 points) What is the effect of the regularisation parameter in LASSO regression on: 1. model complexity; 2. the bias-variance trade off; 3. shrinkage of coefficients; 4. prediction error in the training and test set?

7. (5 points) Assume that your task is to predict the response to a targeted drug for a set of N patients, based on the following available clinical data: expression of the targeted gene, age of the patient, tumour size, height of the patient. Write the cost function of Lasso and Ridge regression. What would you expect to be the effect of these estimators on the estimated coefficients of the four features? Motivate your answer.

8. Backpropagation.
   (a) (1 point) What is the advantage of the backpropagation algorithm over alternative methods for computing the gradient of a deep neural network with respect to its parameters.
   (b) (4 points) The figure below shows a layered view of a neural network with two layers (in addition to the input and loss layers). Describe how the gradient of the loss J with respect to the parameters of the first layer $\Theta^{(1)}$ can be computed using backpropagation.

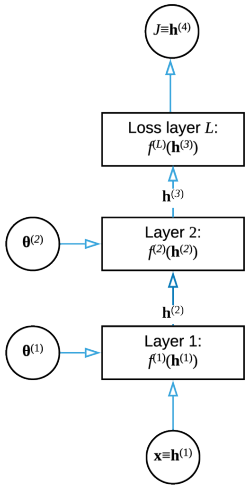

Figure 1: Layered view of a neural network with two layers.

9. (5 points) One way to interpret the workings of a convolutional neural network is to look at saliency maps that show which part of the input image was most predictive for

the output. Name three approaches to obtain a saliency map and briefly describe how each works.

10. (5 points) The figure below shows a generative adversarial network (GAN) with its objective function. Assume that we have a large data set of images with labels describing the content of each image (e.g. normal, benign, malignant). Modify the GAN architecture and the objective function in such a way that it can generate new images from a specific class. Motivate all proposed changes of the architecture and objective function.

$$V^{(D)}(D,G) = \mathop{\mathbb{E}}_{x\sim p_{\text{data}}}[\log D(x)] + \mathop{\mathbb{E}}_{z\sim p_z}[\log(1 - D(G(z)))]$$

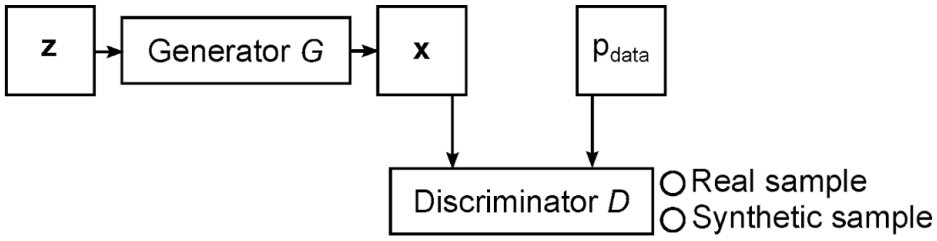

Figure 2: Generative adversarial network model.

11. The following plot shows a classification using support vector classifier.
    (a) (2 points) Mark the five points determining your separating hyperplane.
    (b) (3 points) Consider the point X; what is the effect of moving it to location A and to location B? What are the corresponding values of the slack variable?

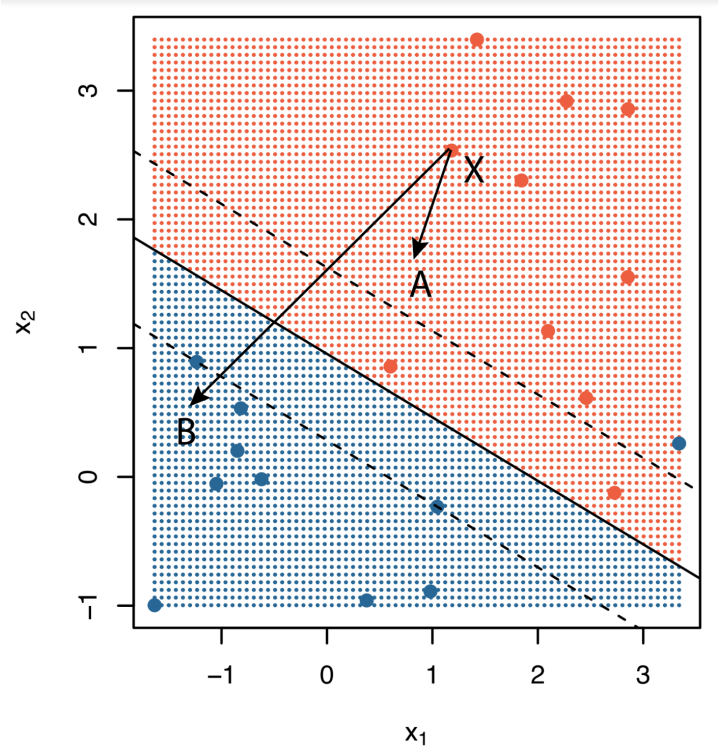

Figure 3: Support vector classifier.

12. (5 points) In which of the two plots below does Principal Component Analysis (PCA) work better and why? On that graph draw a sketch of the principal components.

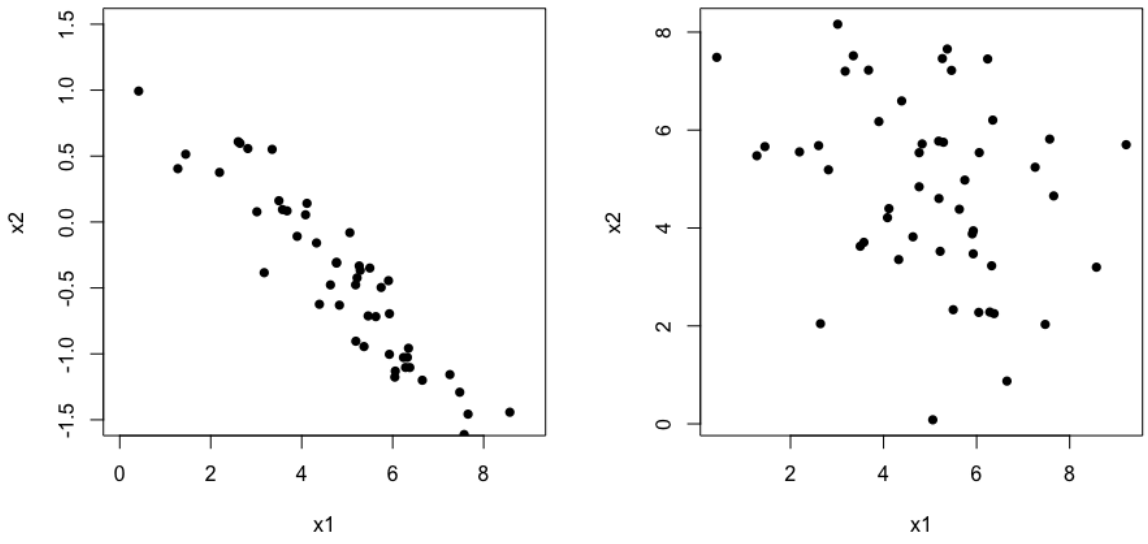

Figure 4: Principal components analysis.

13. (5 points) Write the formulas of the three dissimilarity metrics between two clusters in

hierarchical clustering. In the figure below, draw the dissimilarity between the blue and the red cluster when using single linkage and when using complete linkage. In both cases use Eucledian distance as metric of dissimilarity between two points.

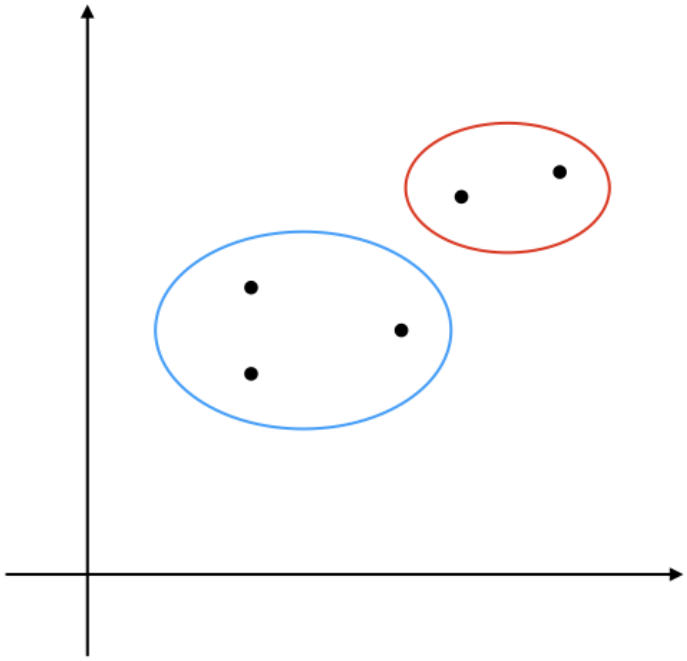

Figure 5: Hierarchical clustering.

14. (5 points) What are the bias and variance properties of each single tree built using: A. Bagging and B. Boosting ? And how do these properties relate to the size of the trees? Motivate your answers.