2. (a) Describe PCA and give its mathematical relation to an eigenvalue decomposition problem. When is it useful to apply PCA and how would you choose the number of principal components to use? [7]

   (b) What is overfitting and how can we prevent overfitting in decision tree models? [6]

   (c) Give the mathematical description of 3 different distance metrics. What distance is suitable for binary-valued vectors? [6]

   (d) Starting from first principles derive the likelihood function for linear regression assuming white noise. Describe the maximisation problem associated with both a Maximum Likelihood and a Maximum-a-Posteriori approach for this setting. [6]