3. (a) Explain the concept of Leave-One-Out-Cross-Validation (LOOCV). When would you use LOOCV? [4]

(b) Describe Multi-dimensional Scaling (MDS) and provide a mathematical and intuitive description of *Stress*. Explain the difference between Metric and Non-Metric MDS. [6]

(c) Give the mathematical description of the Manhattan distance between two vectors $\mathbf{x}_i$ and $\mathbf{x}_j$. What is the relation of the $L1$ norm to the Manhattan distance? [4]

(d) For a specific probabilistic model the probability for an unseen observation $\mathbf{x}^* \in R^D$ to belong to class k is given by:

$$P(t^* = k|\mathbf{X}, \mathbf{t}, \mathbf{x}^*) = \frac{p(\mathbf{x}^*|t^* = k, \mathbf{X}, \mathbf{t})P(t^* = k)}{\sum_{j=1}^{K} p(\mathbf{x}^*|t^* = j, \mathbf{X}, \mathbf{t})P(t^* = j)}$$

i) Is this a generative or a discriminative model? Why? [2]
ii) Re-write the likelihood of this model under the Naive Bayes assumption. [2]
iii) What should $P(t^* = 1)$ be equal to if we have a binary classification problem and 90% of the training observations belong to class 0? [2]

(e) In the context of Bayesian Linear Regression (BLR) the predictive density of interest is given by $p(\mathbf{t}^*|\mathbf{x}^*, \mathbf{X}, \mathbf{t}) = \int p(\mathbf{t}^*|\mathbf{x}^*, \mathbf{w})p(\mathbf{w}|\mathbf{X}, \mathbf{t})d\mathbf{w}$. Apply Bayes rule in the posterior density and describe each resulting term within the context of BLR. What probability density function we typically use for i) the likelihood, and ii) the prior? What type of pdf is the resulting posterior density? [5]