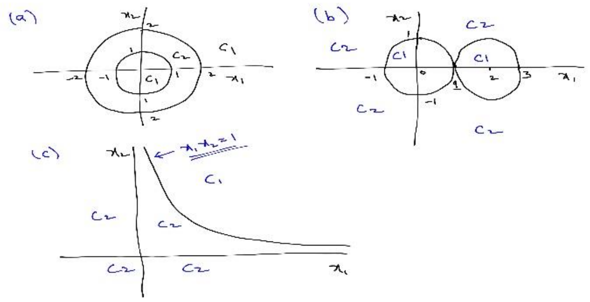
COL 341: Fundamentals of Machine Learning

21st September 2021 Mid-Term Examination Maximum Marks: 100

[30 Marks]

Q1. For each of the binary classification problems given in the figure on the right, use feature engineering (if needed) and find the simplest neural network that can give the desired decision regions. In the figure C1 represents class 1 and C2 represents class 2.



[30 Marks]

Q2. For linear regression problem, consider the following statistical model: (x, y) ~ N(μ, Σ), where $x \in R^m$ represents the m-dimensional feature vector, $y \in R$ represents the target or response variable, the m+1 dimensional vector (X, Y) is distributed according to multi-variate Gaussian distribution with mean $\mu \in R^{m+1}$ and covariance matrix $\Sigma \in R^{m+1 \times m+1}$. To answer this question, you may use the known results about multi-variate Gaussian distributions given in the supplementary sheet.

(a) If the model parameters (μ, Σ) are known then the predictions ($y' \in R$) for the corresponding y for a new example $x \in R^m$, may be done using the maximum likelihood estimator for y. Work out the formula for the MLE estimator of the response variable y for a known value of x.

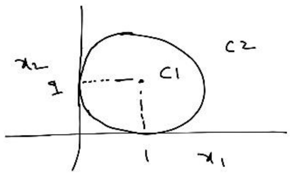
(b) Let us suppose we are given labelled training data corresponding to n examples (X, Y) such that $X \in R^{n \times m}$ and $Y \in R^n$. Find the expression of the log likelihood function L(X, Y; μ, Σ) which represents the log of probability density of observing the given data (X, Y) given the model parameters (μ, Σ).

(c) The model parameters may be learned by using the maximum likelihood estimator (MLE). Work out the closed form expressions for the MLE estimators for the model parameters (μ, Σ).

(d) Now suppose that you are given an example x where some of the feature values are missing. For simplicity of notation, we assume that x_1 .. x_k are known and x_{k+1} .. x_m are missing. Work out the expression for the MLE estimator of y using only the non-missing values x_1 .. x_k. Also give a MLE estimator for estimating the missing values x_{k+1} .. x_m.

[40 Marks]

Q3. Consider the binary classification problem given in the figure on the right.



(a) Will logistic regression ever be able to learn the correct decision boundaries? Why?

(b) Show that by using suitable feature engineering logistic regression may be able to learn the correct decision boundaries. Give complete details of your proposed feature engineering method.

(c) Consider the following latent variable model for binary classification problems:

$$y_i = \begin{matrix} 1 & if\ w'x_i + \epsilon > 0 \\ 0 & otherwise \end{matrix}$$

where y_i is the binary response variable (or target variable) and x_i is the m-dimensional feature vector. ϵ is the latent (unobserved) error which is distributed by the standard logistic distribution. The probability density function of the standard logistic distribution is given by: $f(\epsilon) = \frac{e^{-\epsilon}}{(1+ e^{-\epsilon})^2}$. The vector w is m dimensional vector of model parameters to be learned (here w' represents the transpose of w). Give n training examples (X, Y) such that $X \in R^{n \times m}$ and $Y \in \{0, 1\}^n$. Find the log likelihood function L(X, Y; w) for this model.

(d) To learn the model parameters, we need to do gradient descent on the log likelihood function. Find the gradient of the log likelihood function with respect to the model parameters. Write the simplified expression of the gradient.