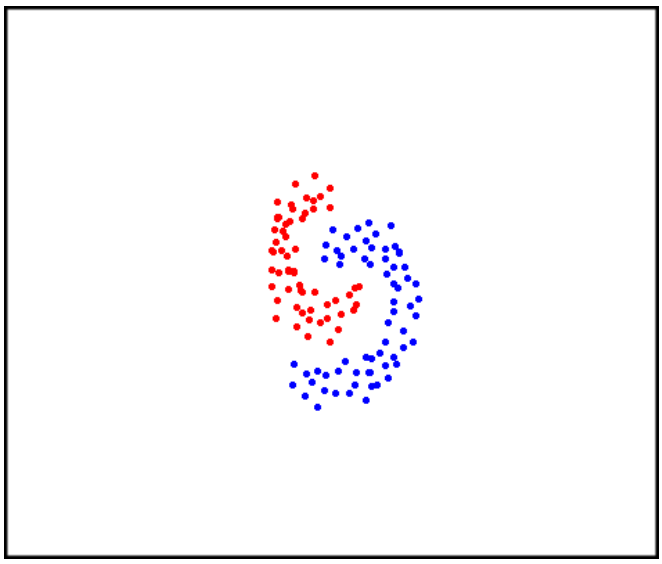
Question 1: Probabilities: ..... $(\Sigma = 2)$

(a) Given $p(x|c)$ and $p(c)$ for $c \in \{1, \ldots, C\}$, give an equation to compute $p(c|x)$. **(2 pts)**

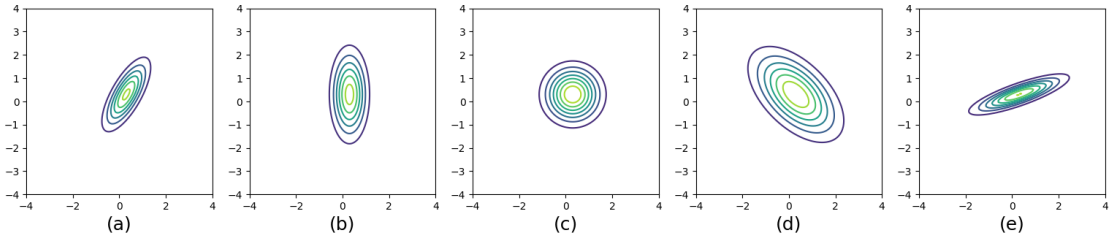
Question 2: Probability Density Estimation: ..... $(\Sigma = 4)$

(a) You are given the set of data points $\{\mathbf{x}_1, ..., \mathbf{x}_N\}$ with $\mathbf{x}_i \in \mathbb{R}^2$ as shown in the plot below. Each point belongs to one of two classes as denoted by blue and red colors. You decide to use the K-means algorithm with $K = 2$ to cluster the data in order to classify the points. Do you think that the result will be accurate? Give a justification for your answer. **(1 pt)**



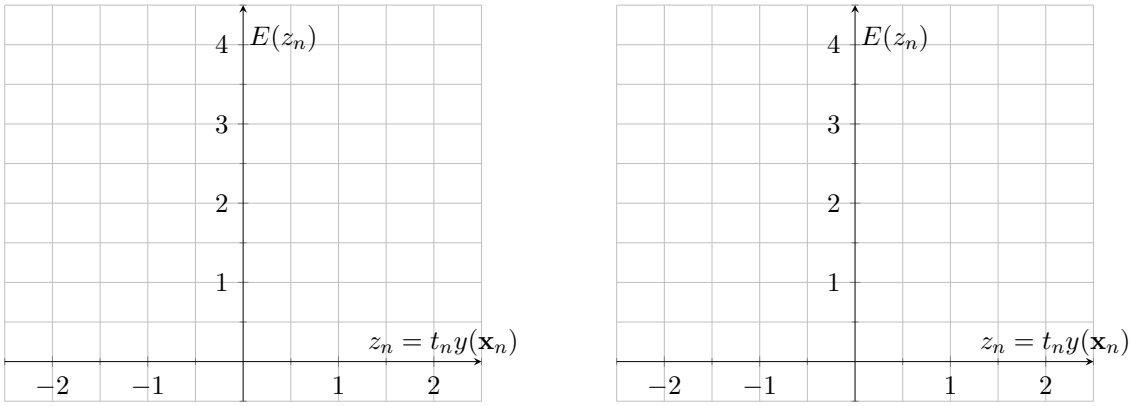
(b) What is a good strategy for initializing the parameters of a Mixture Model in Expectation-Maximization? **(1 pt)**

(c) Which of the images below corresponds to the contour plots of a Gaussian with covariance matrix $\begin{bmatrix} 0.2 & 0 \\ 0 & 1.2 \end{bmatrix}$? **Exactly one answer is correct. Justify your choice. Answers with no justification will not receive any points.** **(2 pts)**



Question 3: Linear Discriminants: ..... $(\mathbf{\Sigma = 3})$

(a) Plot the error contribution plot for the squared error function for the case of binary classification (i.e. target values $t_n \in \{-1, 1\}$). **(1 pt)**



Hint: Use the second plot to correct your drawing. Cross out wrong answers.

(b) In their basic form, linear discriminants are given by $y(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b$. Write down the error function that is optimized by a linear least-squares classifier. Explain the terms and variables you use. **(2 pts)**