Problem 1: Recall the naive Bayes model.

- For simplicity, we only consider binary features

$$x_i \in \{0,1\}, \quad \text{i.e.,} \quad \boldsymbol{x} \in \{0,1\}^d$$

- The generation model is

$$t \sim \text{Categorical}(\pi_1, \cdots, \pi_K)$$

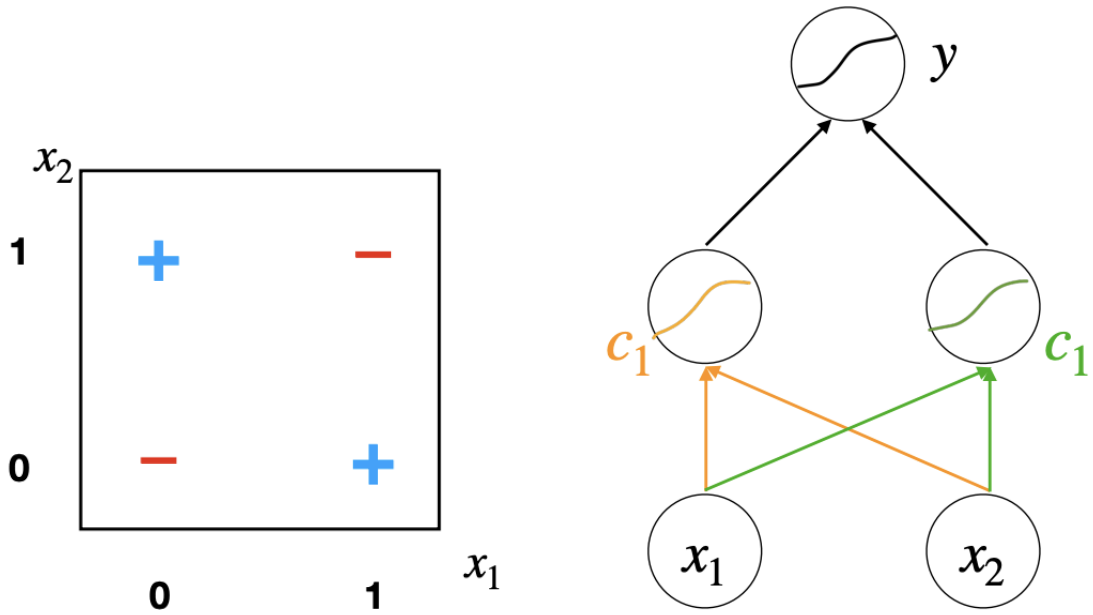
$$\boldsymbol{x}_i | t = k \sim \text{Bernoulli}(p_{k,i})$$

  Here: A Bernoulli distribution parametrized by π means that
  $\Pr[X = 1] = \pi$ and $\Pr[X = 0] = 1 - \pi$.
  It is a special case of categorical distributions in that only two cases are considered.

- Such a model can be used to represent a document in text classification. For example, the target indicates Spam or NotSpam. The feature indicates if a word in the vocabulary occurs in the document.

Show that the decision boundary of naive Bayes is also linear.

Problem 2: Give a set of parameters of the below stack of logistic regression models to accomplish the non-linear classification.



Problem 3: Using the above example to show the optimization of neural networks is non-convex.

Hint: As mentioned in class, you can rename c_1 as c_2 and c_2 as c_1. Then, you will have two models. Interpolating them will give you a very bad model.

END OF W9