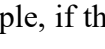
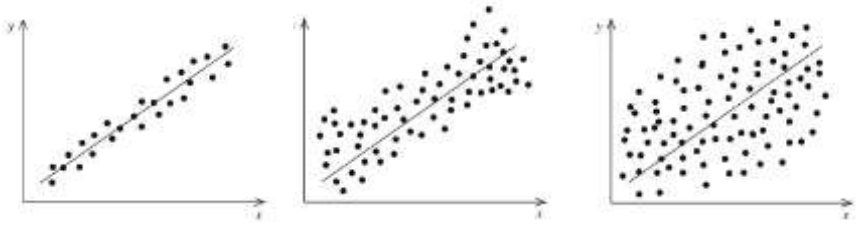
HW #2 Due: 3/22/2023

1. We mentioned the HIV detection problem in the Bayesian decision theory. Use the Bayes theorem to confirm the given answer (i.e., 9%). To answer this problem, you need to distinguish two different conditions:
   - False positive is a conditional probability *P*(reagent is negative | patient is infected). Same argument for false negative.
   - When a patient is given a positive test result, it is actually *P*(patient is infected | reagent is positive)
2. We explained how to use Naïve Bayesian classifier to classify colored squares and circles in the lecture. Following the example, if the following pattern  is added to the training set, redo the classification of ■.
3. Below are scatterplots of Gaussian random points. Among these three plots, which one has the smallest correlation coefficient, and which one has the largest correlation coefficient? Is there any plot corresponding to negative correlation coefficient? Explain your answers.



(a) (b) (c)

4. If a sequence of coin tossing has the results of H,H,T,T,H,H,H, follow the MAP method given in the ppt file to **numerically** (i.e., from the beta distribution) find θ_{MAP} based on a likelihood plot for $a = b = 5$. You need to include you program to show how the likelihood plot is drawn.
5. In this problem, you are asked to use the Naïve Bayes classifier to classify the Breast Cancer Wisconsin (Original) Data Set, directly from the sklearn or downloading from https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Original%29). To simplify the problem, we just want to keep 8 attributes out of 9. (hint: one attribute is useless. Which one is it?) To begin one trial, randomly draw 70 % of the instances from each class for training, and the rest 30% for testing to obtain the accuracy. Repeat the trials 10 times to report the average accuracy. Note: this dataset has missing attributes. **Explain how you handle missing attributes.**