Pattern Recognition Experiment 2

EECS26 111060024 蔡孟伶

1. Generate two data set X_1 and X_2, each with $N = 1{,}000$ 3-D vectors, for the two classes C_1 and C_2, respectively. Assume C_1 and C_2 are modeled by normal distributions with the parameters: $\boldsymbol{\mu}_1 = [0,0,0]^T$, $\boldsymbol{\mu}_2 = [1,5,-3]^T$, $\Sigma_1 = diag(3,5,2)$, and $\Sigma_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 1 \\ 0 & 1 & 6 \end{bmatrix}$
   - (a) Write the code to find the maximum-likelihood values $\hat{\mu}_{ML}$ and $\hat{\sigma}^2{}_{ML}$ *individually* for each of the three features x_i $(i = 1,2,3)$ for the class C_1.
   - (b) Write the code to find the maximum-likelihood estimations $\hat{\boldsymbol{\mu}}_{ML}$ and $\hat{\boldsymbol{\Sigma}}_{ML}$ to the three-dimensional Gaussian data $\mathbf{x} = [x_1, x_2, x_3]^T$ for C_1.
   - (c) Compare your results for the mean and variance of each feature calculated in (a) with those in (b). Explain why they are the same or different.
   - (d) Assume the three features are uncorrelated, i.e., $\boldsymbol{\Sigma} = diag(\sigma_1^2, \sigma_2^2, \sigma_3^2)$. Write the code to find the maximum-likelihood estimation of $\hat{\boldsymbol{\mu}}_{ML}$ and $\hat{\boldsymbol{\Sigma}}_{ML}$ for C_2.
   - (e) Apply your code in (b) to find $\hat{\boldsymbol{\mu}}_{ML}$ and $\hat{\boldsymbol{\Sigma}}_{ML}$ for C_2. Compare your results for the variances obtained in (d) with those in (b). Explain why they are the same or different.

(c) The mean estimator derived in (a) and (b) are the same. Additionally, the variance calculated at (a) is the same as the covariance matrix's diagonal line. I think it is reasonable that sigma(i, j) means the variance between feature i and feature j. Therefore, the diagonal line is the variance within the individual class. They are both calculated under biased variance.

(e) Inside (d), assuming the features were uncorrelated, then the mean estimator and variance estimator would be the same as (a). Therefore, this is another case of comparing the result derived from (a) and (b). The variance value will be the same, but for the dataset of C2, the covariance matrix of generating the dataset is not like what (d) assume. Therefore, although they have the same value for individual class variance, the covariance matrix is not the same. (d) has 0 on sigma(i, j), i != j. (e) computes the between feature covariance, which is not 0.

2. Write the code that generates the Bernoulli samples $X = \{x_1, \ldots, x_N\}, x_i \in \{0,1\}$ with (a) $p = 0.7, N = 1000$, and (b) $p = 0.7, N = 5000$; and the code that calculates the estimate $\hat{p}_{ML}$ from the sample X. Compare your estimated $\hat{p}_{ML}$ with the ground truth $p = 0.7$ for the two data sets and draw your conclusions.

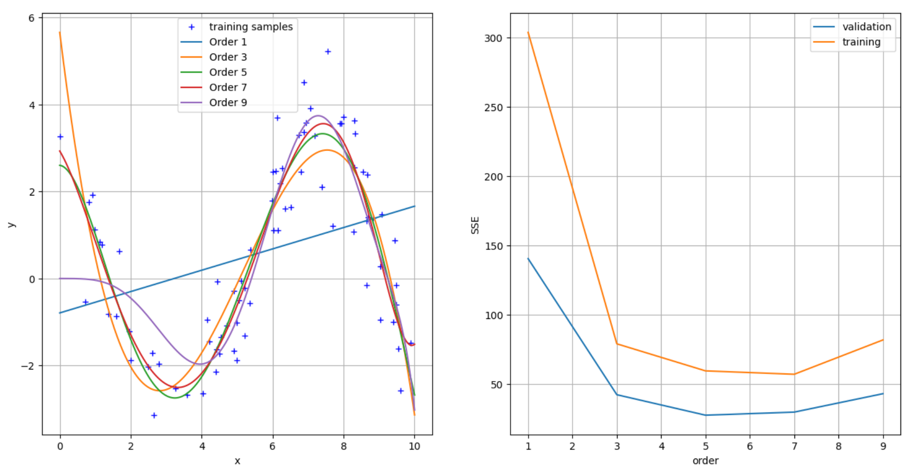
Conclusion:

Since the p ML estimator is equal to the mean of the sample value, by testing a few times, (b) usually has estimated p ML closer to 0.7 than (a), and the reason is likely because that (b) has more samples, and could generate dataset that fits better to the ground truth we set.

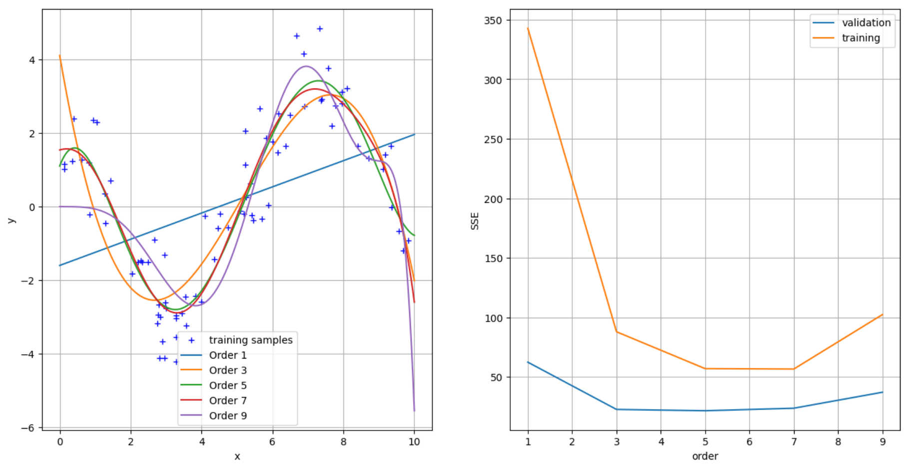
3. Write the code that generates 100 samples $X = \{(x_i, y_i), i = 1, \ldots, 100\}$ ten times by $y = 3\sin(0.8x + 2) + \varepsilon,\ \varepsilon \sim N(0,1)$ for $x \in [0,10]$. Divide your samples into two as training and validation sets.
   (a) Use a polynomial in x of order k $(k = 1, 3, 5, 7, 9)$ to fit the training data. Refer to Fig 4.5 (in Ch4's slides) to show your training samples and the fitted curves.
   (b) Compute error on the validation set. Refer to Fig. 4.7(b) to show the regression error vs. polynomial order.

The following is the result figure of (a) and (b) for each dataset, I think usually that taking order = 5 would get low SSE error than other choices inside cross validation.

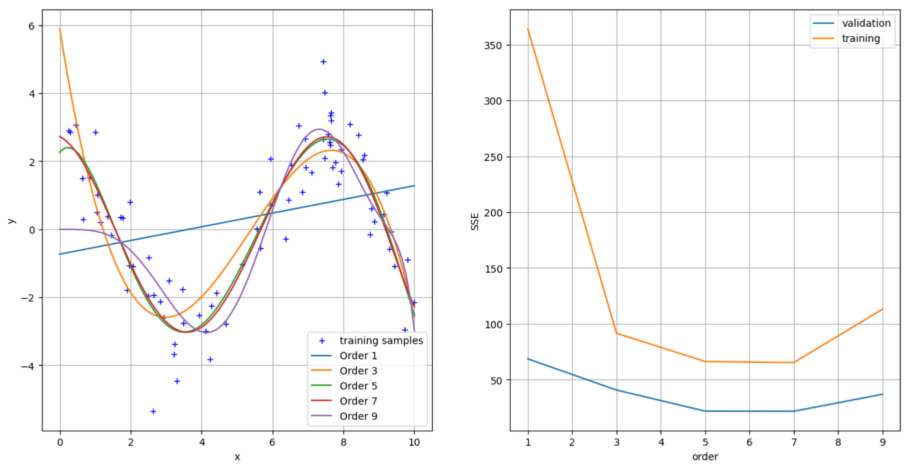
Dataset 0



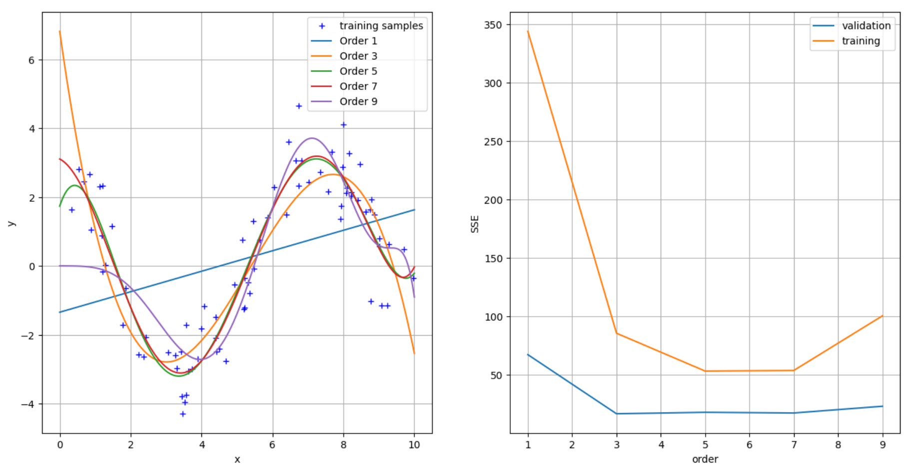
Dataset 1



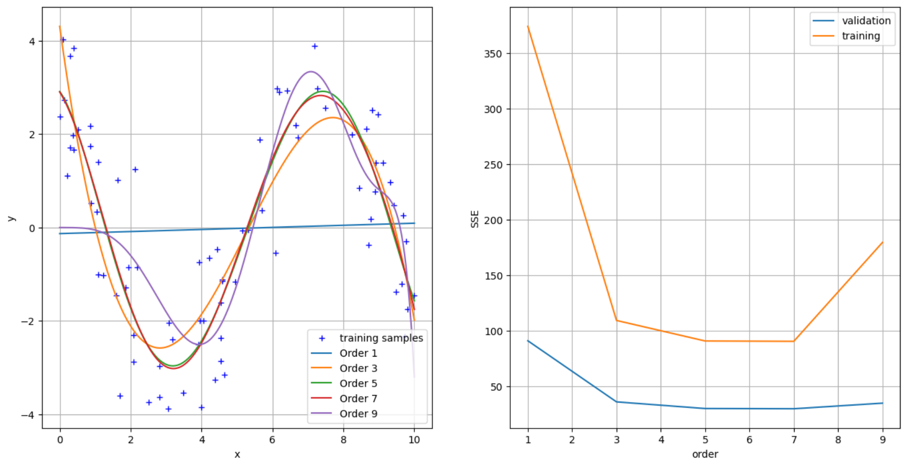
Dataset 2



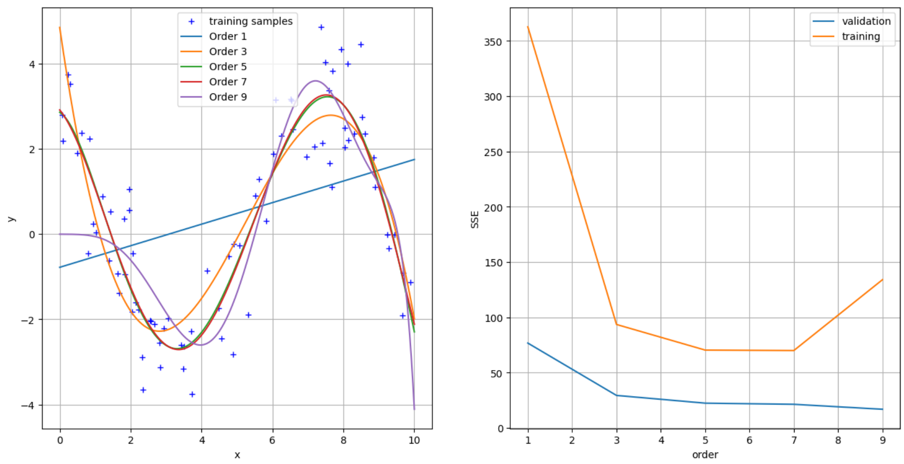
Dataset 3



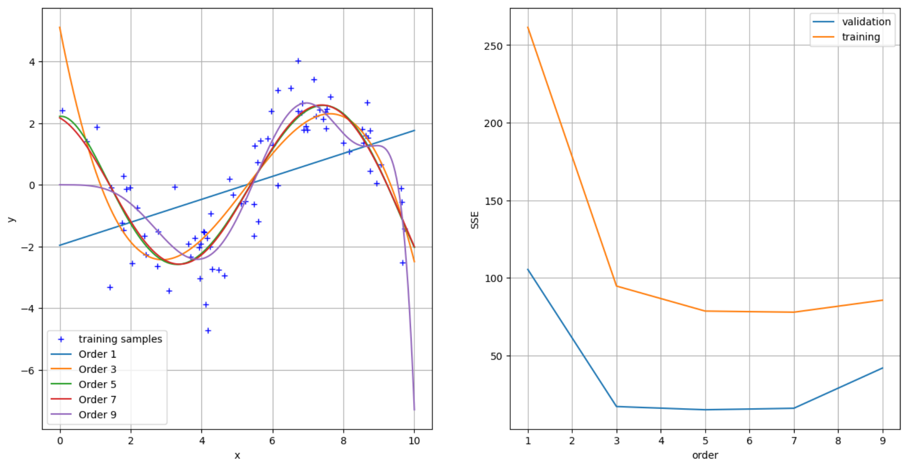
Dataset 4



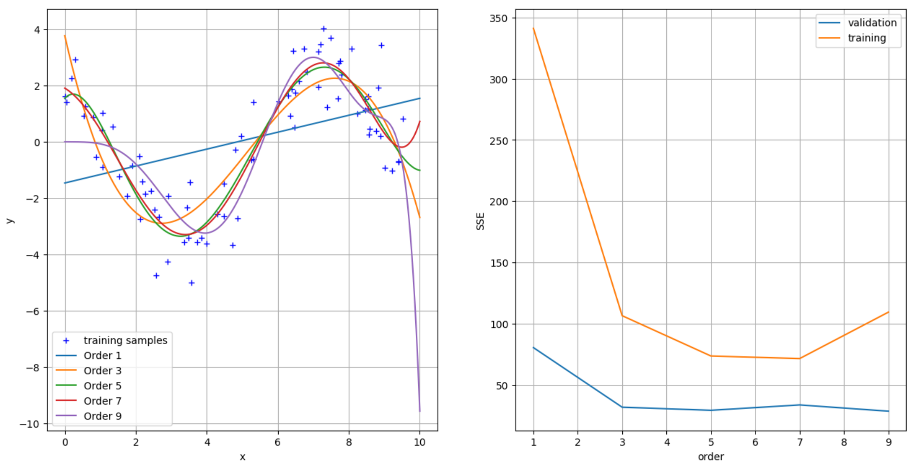
Dataset 5



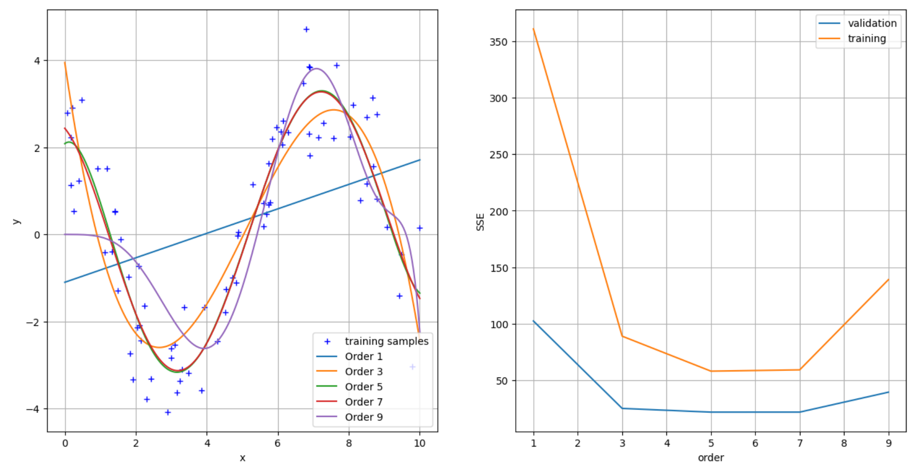
Dataset 6



Dataset 7



Dataset 8



Dataset 9