1. Solve the regression problem using the dataset given below. Assume a linear model.

|  |  |  |
| --- | --- | --- |
| Pollution level | area | Price of house |
| -1 | 13 | -52.7 |
| 0 | 13.5 | -52.1 |
| 1 | 14 | -51 |
| 2 | 14.5 | -49.5 |
| 3 | 15 | -49.3 |
| 4 | 15.5 | -48.1 |
| 5 | 16 | -46.8 |
| 6 | 16.5 | -46.2 |
| 7 | 17 | -45 |

1. How will you know whether the model you have trained is underfitted or overfitted?
2. Plot testing error as a function of complexity of model.
3. Write a short note on bias-variance tradeoff.
4. From the dataset given below, answer the following questions:

(a) dimension of the problem, (b) sample size, (c) how many discriminant functions do you need to construct? (d) How many parameters do you need to estimate if you want to solve the problem using discriminant functions? (Answer without numerically solving the problem).

|  |  |  |
| --- | --- | --- |
| Annual income | age | Type of car purchased |
| 10 | 30 | I |
| 15 | 35 | II |
| 10 | 50 | I |
| 12 | 39 | I |
| 20 | 25 | III |
| 22 | 30 | II |
| 30 | 45 | III |
| 32 | 21 | I |
| 49 | 49 | II |
| 9 | 45 | II |
| 11 | 32 | III |
| 43 | 32 | II |
| 21 | 300 | III |
| 26 | 20 | II |
| 50 | 25 | I |
| 60 | 20 | I |
| 22 | 68 | III |

1. Consider the following fish classification problem assuming likelihood functions Normally distributed. Find the equation of the decision boundary.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Length of fish in cm | 20 | 22 | 24 | 26 | 28 | 30 |
| Seabass | 11 | 9 | 8 | 7 | 3 | 2 |
| Salmon | 3 | 5 | 6 | 9 | 8 | 9 |

There are 11 Seabass of length 20 cm, 3 Salmon of length 20 cm, 9 Seabass of length 22 cm, and so on.

1. Define VC dimension.
2. Which loss functions do you typically use for a regression, logistic regression, and linear discriminant analysis problem ?