Exercise: Generalization

Links:

- Task1: Multiple choice
- Task2: Questions

Task1: Multiple choice

Question 1.1

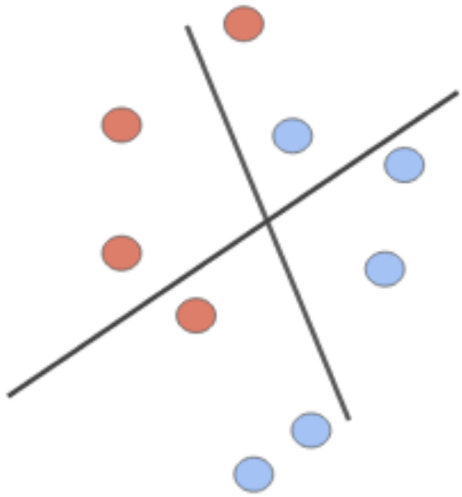
What does this represent?

Options:

1. A model
2. An example
3. A hypothesis
4. In-sample

Question 1.2

What is an effect of choosing the best (in-smaple) from two hypothesis?



Options:

1. Out-of-sampe error decrease
2. Lower probability of generalization
3. In-sample error increase
4. Higher probability of generalization

Question 1.3

Which statement is true?

$$\hat{y} = sign(ax^3 + bx + c)$$
$$\hat{y} = sign(ax + c)$$

Options:

1. $ax^3 + bx + c$ have a larger hypothesis space
2. $ax^3 + bx + c$ have a smaller hypothesis space
3. Their hypothesis space is of the same size

Question 1.4

Which statement is true?

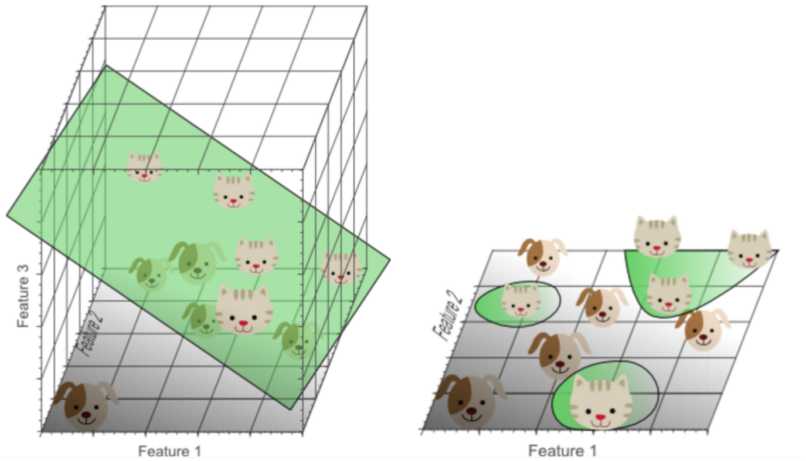
$$\hat{y} = sign(ax^2 + c)$$
$$\hat{y} = sign(ax + c)$$

Options:

1. $ax^2 + c$ have a larger hypothesis space
2. $ax^2 + c$ have a smaller hypothesis space
3. Their hypothesis space is of the same size

Question 1.5

What is the VC dimension of a linear classifier in 3D (plane)



Options:

1. $d_{VC} = 3$
2. $d_{VC} = 4$
3. $d_{VC} = 7$
4. $d_{VC} = 8$

Question 1.6

What is the VC dimension of a n dimensional linear classifier?

Options:

1. $n + 1$
2. $n^2 + 1$
3. $n^2 - (n-1)^2 - 1$

Question 1.7

What is true about the VC dimension of a model?

Options:

1. A model with high VC dimension is more likely to underfit
2. More training examples will give a lower VC dimension
3. A higher VC dimension give a better classifier
4. A model with high VC dimension is more likely to overfit

Question 1.8

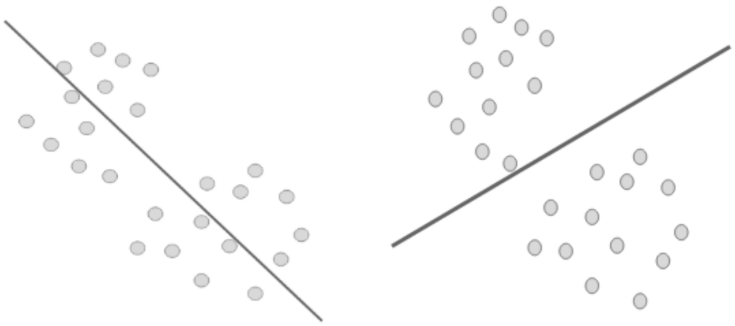
Will using the K-Nearest Neighbors classifier with k=1 imply:

Options:

1. Most likely to overfit
2. Most likely to underfit
3. Depends on the data

Question 1.9

Given an unknown class distribution, which line is probably the best classification boundary?

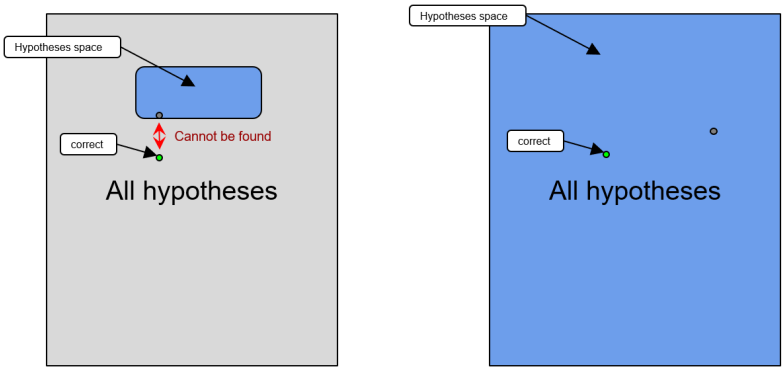


Options:

1. The left line
2. The right line
3. They are both equally good

Question 1.10

What effect does adding a regularization term has on the hypothesis space?



Options:

1. No effect
2. Decreasing the hypothesis space
3. Increasing the hypothesis space

Task2: Questions

Give short answers to the following questions.

Question 2.1

Early stopping is often used to prevent (limit) overfitting. Is the out-of-sample error estimate from the validation set a good estimate?

Answer:

Question 2.2

How can we improve the out-of-sample error when we have a small dataset only?

Answer: