*What are the **advantages** of a nearest neighbour classifier?*

1

*What are the **disadvantages** of a nearest neighbour classifier?*

2

What is the most important concept in machine learning?

3

What are the three 'ingredients' of a machine learning algorithm?

4

What does this equation calculate?

$$a = \sum_{i=1}^{F} x_i w_i$$

5

What is the perceptron learning rule?

6

What does this equation calculate?

$$a = \frac{1}{1 + exp(-\sum_{i=1}^{d} w_i x_i)}$$

7

Decision trees are good at handling _____ *data but worse at handling* _____ *data.*

8

- Very computationally expensive for every classification
- Complexity depends on the number of dimensions

2

- Very accurate
- No learning process

1

The model, the error function and the learning algorithm.

4

Never assume that you have all the data.

3

$$newWeight = oldWeight + 0.1 \times (trueLabel - output) \times input$$

6

The activation of the perceptron

5

Decision trees are good at handling categorical data but worse at handling continuous data.

8

The activation of the perceptron for non-linearly separable data

7

What does this equation calculate?

$$H(X) = -\sum_i p(x_i) \log_2 p(x_i)$$

9

The 'information' contained in a varibale is called the _____.

10

Explain the process of cross validation.

11

What factors should affect our decision on the best value of k?

12

What is the ensemble approach to machine learning?

13

Briefly describe bootstrapping

14

On average, what is the percentage of data points that are left unselected?

15

Explain bagging

16

The 'information' contained in a varibale is called the entropy.

10

The entropy of a variable X

9

1. *Accuracy*
2. *Training time and space complexity*
3. *Testing time and space complexity*
4. *Interpretability*

12

1. *Break the data evenly into N chunks*
2. *Leave one chunk out*
3. *Train on the remaining $N-1$ chunks*
4. *Test on the chunk you leave out*
5. *Repeat until all chunks have been used to test*
6. *Plot the average and error bars for the N chunks*

11

Bootstrapping is the process of generating multiple data sets from an original.

14

Select a class of models, fit multiple models to training data (called base learners), use the models as a committee to vote on testing data.

13

Generate m bootstraps and train a model on each one. When the testing data arrives a simple majority vote takes place.

16

36.8%

15

Explain boosting

17

What type of classifier models a classification rule directly and models the probability of class memberships based on input data?

18

What type of classifier makes a probabilistic model of data within each class?

19

What type of classifier uses probabilities to classify data?

20

What is the formula to work out $P(c|X')$ Where c is a class and X' is an example?

21

What is the formula to work out a Gaussian model?

22

What are the two data representation methods that we talk about in clustering analysis?

23

What is the formula to work out Minowski distance.

24

A discriminative classifier

18

Get a data set, take a bootsrap and train a model on it. See which examples the model got wrong then upweight those 'hard' examples and downweight the 'easy' ones. Now go back to training a model, but now you have a weighted bootstrap.

17

A probabilistic classifier

20

A generative classifier

19

$$\frac{1}{\sigma\sqrt{2\pi}}exp(-\frac{(x-\mu)^2}{2\times\sigma^2})$$

22

$$P(c|X') = [P(x_1|c)P(x_2|c)....P(x_n|c)]P(c)$$

Where x is a feature in the example.

21

$$d(x,y) = \sqrt[p]{(x_1-y_1)^p + (x_2-y_2)^p ... + (x_n-y_n)^p}$$

24

Data matrices and distance matrices.

23

What is the formula for Manhattan distance?

25

What is the formula for Euclidean distance?

26

What is the cosine measure equation

27

What is the formula for the distance between symmetric binary attributes?

28

What is the formula for the distance between asymmetric binary attributes?

29

$$d(x,y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 ... + (x_n - y_n)^2}$$

26

$$d(x,y) = (x_1 - y_1) + (x_2 - y_2) + (x_n - y_n)$$

25

$$d(x,y) = \frac{b+c}{a+b+c+d}$$

28

$$\frac{x_1y_1 + ... + x_ny_n}{\sqrt{x_1^2 + ... + x_n^2}\sqrt{y_1^2 + ... + y_n^2}}$$

27

$$d(x,y) = \frac{b+c}{a+b+c}$$

29