**Which ML method can be described as a combination of branches and leaves?**

====

Linear Regression

====

#Decision Tree

====

Support Vector Machine

====

Neural Networks

++++

**In which ML method we use the concept of the “widest possible road”?**

====

Linear Regression

====

Decision Tree

====

#Support Vector Machine

====

Neural Networks

++++

**Which of the listed ML methods is the oldest one?**

====

#Linear Regression

====

Decision Tree

====

Support Vector Machine

====

Neural Networks

++++

**Which of the listed ML methods was designed after the biological prototype?**

====

Linear Regression

====

Decision Tree

====

Support Vector Machine

====

#Neural Networks

++++

**Can the Linear Regression method be applied to non-linear data?**

====

No

====

#Yes

++++

**If you find that your dataset demonstrates a strong multicollinearity, what will you do in this case?**

====

#Remove the variable

====

Add more variables

====

Join two variable into one

====

Choose a random subset with replacement

++++

**Which of the listed equations are linear (multi-choice question):**

====

#y = ax + b

====

#Ax + By + C = 0

====

#Ax + By + Cz + D = 0

====

#Ax<sub>1</sub> + Bx<sub>2</sub> = 0

====

Ax + By + Cxy + Dx<sup>2</sup> + Ey<sup>2</sup> = 0

++++

**I have collected a data about the shop’s customers, such as their age, gender, payment methods, product categories, quantity, price, etc. Now I want to analyze them and separated into some meaningful groups. What ML methods I will use in this case?**

====

#Clusterization

====

Classification

====

Regression

++++

**I am managing an online shop, and I have recorded how many users visit my website every day over one year. Now I need to predict how many visitors will attend it next Friday. Which ML method will I use for this task?**

====

Classification

====

#Regression

====

Association

====

Dimensionality Reduction

++++

**The online shop wants to organized a big sell of the women clothes. We know what our customers buy, and we need to find who of them are men and who are women. Which method will allow us to separate them into two groups? (multi-choice question)**

====

#Decision Tree

====

#Decision Forest

====

#Support Vector Machine

====

#Neural Networks

====

#k-Nearest Neighbor

++++

**(Class attendance check) What item is bought often together with bread and diapers?**

====

#beer

====

milk

====

coca-cola

====

eggs

++++

**When building a classification Decision Tree, we split the data based on a simple question about one of the attributes. But how do we choose which attribute to use?**

====

#by finding which one reduces Entropy the most

====

by finding which one increases Entropy the least

====

by finding the average Entropy

====

by finding the minimal MSE

++++

**We are creating a Decision Tree model predicting the stock prices, which criterion is used to find the best split of the data?**

====

#Mean squared error (MSE)

====

Information gain

====

Entropy

++++

**How many points are needed to build the minimal SVM model?**

====

1

====

#3

====

5

====

10

++++

**Kernels in SVM models are used with the goal of: (multi-choice question)**

====

#building a non-linear model

====

building a linear model

====

#improving an underfitted model

====

improving an overfitted model

++++

**Support Vector Machine is tuned by setting the regularization parameters (C) responsible for soft (low C) or hard (high C) margins. On reducing C, the model becomes: (multi-choice question)**

====

#more resistant to the outliers

====

less resistant to the noise

====

#more stable

====

less general

++++

**SVM with the Gaussian kernel is managed by gamma (γ). High γ leads to the tight borders, best embracing the training points. Which value of γ leads to the underfitting?**

====

#low value

====

high value

++++

**Which of the ML methods are deterministic? (multi-choice question)**

====

#Decision Tree

====

#Support Vector Machine

====

Neural Networks

++++

**Which of the ML methods are stochastic? (multi-choice question)**

====

#Decision Forest

====

Support Vector Machine

====

#Neural Networks

++++

**Can a SVM model be used for multi-class classification?**

====

no, only binary classification

====

#yes, but need to make an ensemble

====

yes, naturally supports multi-class

++++

**If SVM model in classification is “the widest road”, then in regression it should be:**

====

#the thinnest bridge

====

the straightest road

====

the lowest bridge

++++

**Which of the listed problems belongs to the classification task? (multi-choice question)**

====

#sentiment analysis of texts: positive, negative, or neutral

====

#spam detection

====

#image recognition

====

#fraud detection

====

stock prediction

++++

**In what aspect, Deep Neural Networks are different from the classic ANN?**

====

#in the number of the hidden layers

====

in the number of the output layers

====

in the links connecting neurons inside the layer

====

in their ability to connect a neuron to several layers simultaneously

++++

**The model of the simple perceptron contains bias, what is it?**

====

#constant

====

linear function

====

non-linear function

====

weight of the link

++++

**Which of the following statements is true about the gradient? (multi-choice question)**

====

#the gradient is the derivative

====

the gradient is a vector directed toward a minimum

====

#the gradient has magnitude

++++

**A Neural Network model is used to classify the shop customers into three groups: girls, middle age women, and elderly women. How many ANN models are needed in the simplest case to find all three groups?**

====

#1

====

2 (ensemble)

====

3 (ensemble)

++++

**You create a regression model. Which model quality indicator will you use?**

====

#Mean Square Error (MSE)

====

accuracy

====

precision

====

recall

====

F1

++++

**Which of the model quality indicators in the list are reciprocal? (multi-choice question)**

====

Mean Square Error (MSE)

====

accuracy

====

#precision

====

#recall

====

F1

++++

**A customer calls to the bank support service and complains: “I didn’t make this transaction, who stole my money?” To which cell of the confusion matrix would you count this call?**

====

True Positive

====

True Negative

====

False Positive

====

#False Negative

++++

**A customer calls to the bank support service and complains: “Why is my transaction blocked? I need to buy this asap.” To which cell of the confusion matrix would you count this call?**

====

True Positive

====

True Negative

====

#False Positive

====

False Negative

++++

**A man went to the hospital to make the blood test. On the next day, he receives the result saying that he is pregnant. To which cell of the confusion matrix would you count this result?**

====

True Positive

====

True Negative

====

#False Positive

====

False Negative

++++

**In the binary classification, we use a so-called baseline (random) model. What is the theoretical value of the precision of such a model?**

====

10%

====

25%

====

#50%

====

100%

++++

**We create an ensemble of three models, one DT and two SVM, followed by an ANN meta-model. Which kind of ensemble technics do we use?**

====

#blending

====

bagging

====

boosting

++++

**We create an ensemble of three SVM models. Which kind of ensemble technics do we use?**

====

blending

====

#bagging

====

boosting

++++

**We create an ensemble of 40 DT models, consecutively correcting the previous model. Which kind of ensemble technics do we use?**

====

blending

====

bagging

====

#boosting

++++

**There are two models with accuracy 0.8 and 75%, respectively. Which model is better?**

====

#model 1

====

model 2

++++

**There are two models with accuracy 0.6 and 0.4, respectively. Which model is better?**

====

#model 1

====

model 2

++++

**There are two models with precision 0.8 and 0.2, respectively. Which model is better?**

====

#model 1

====

model 2

++++

**There are two models with precision 0.67 and 0.66, respectively. Which model is better?**

====

#model 1

====

model 2

++++

**There are two models with recall 0.7 and 0.6, respectively. Which model is better?**

====

#model 1

====

model 2

++++

**There are two models with recall 0.81 and 0.80, respectively. Which model is better?**

====

#model 1

====

model 2

++++

**There are two models with F1 measure 0.7 and 0.3, respectively. Which model is better?**

====

#model 1

====

model 2

++++

**There are two models with F1 measure 0.7 and 0.6, respectively. Which model is better?**

====

#model 1

====

model 2

++++

**What is the theoretical lowest value of the accuracy of a model doing binary classification?**

====

0

====

0.25

====

#50%

====

0.75

++++

**What is the theoretical lowest value of the precision of a model doing binary classification?**

====

0

====

0.25

====

#50%

====

0.75

++++

**What is the theoretical lowest value of the recall of a model doing binary classification?**

====

0

====

0.25

====

#50%

====

0.75

++++

**Which model quality measures are called ‘reciprocal’? (multi-choice question)**

====

#recall

====

#precision

====

accuracy

====

F1

====

AUC

++++

**Why are cancer tests in medicine often made in two steps?**

====

#to save money

====

for better precision

====

to avoid mistakes

++++

**A pregnant woman went to the hospital to make the pregnancy test. On the next day, she receives the result saying that she is pregnant. To which cell of the confusion matrix would you count this result?**

====

#True Positive

====

True Negative

====

False Positive

====

False Negative

++++

**A patient went to the hospital to take an x-ray of his tooth. On the next day, a doctor calls him and says that the patient is pregnant. To which cell of the confusion matrix would you count this call?**

====

True Positive

====

True Negative

====

#False Positive

====

False Negative