Which of these are possible loss functions ?

Veuillez choisir au moins une réponse :

- ☑ a. Cross-entropy
- ☐ b. Softmax
- ☑ c. Relu
- ☑ d. Square distance
- ☐ e. Sigmoid

What does a neuron compute ?

Veuillez choisir au moins une réponse :

- ☑ a. A neuron computes a linear function (z = Wx + b) followed by an activation function
- ☐ b. A neuron computes an activation function followed by a linear function (z = Wx + b)
- ☐ c. A neuron computes the mean of all features before applying the output to an activation function
- ☐ d. A neuron computes a function g that scales the input x linearly (Wx + b)

With Adaboost :

Veuillez choisir au moins une réponse :

- ☐ a. Adaboost never overfits
- ☑ b. the training error can reach an optimal point
- ☐ c. the training error can be less than the Bayes risk

Common criteria for learning decision (classification) trees are :

Veuillez choisir au moins une réponse :

- ☑ a. mutual information
- ☑ b. Gini criterion
- ☐ c. hinge loss
- ☐ d. square loss

Universal approximators are hypothesis space which are dense in the space of continuous functions :

Veuillez choisir au moins une réponse :

- ☐ a. all SVM with nonlinear kernels are universal approximators
- ☑ b. the family of regression trees with a fixed size is not a universal approximator
- ☐ c. any linear regression model is a universal approximator

Which of the following can be used to regularize a neural network ?

Veuillez choisir au moins une réponse :

- ☑ a. Early stopping
- ☑ b. Weight decay
- ☑ c. Dropout

About decision trees :

Veuillez choisir au moins une réponse :

- ☑ a. A regression tree computes a piece-wise constant function
- ☐ b. A decision tree globally maximizes a mutual information criterion
- ☐ c. None of the other answers are correct
- ☐ d. A binary decision tree only applies to binary classification

The complexity (richness of the family) of a learned SVM :

Veuillez choisir au moins une réponse :

- ☑ a. depends on the value of hyperparameter C (the one before the sum of errors)
- ☐ b. depends on the number of Support Vectors
- ☑ c. depends on the kernel definition

Tackling a supervised learning problem, when should I combine the predictions of several models :

Veuillez choisir au moins une réponse :

- ☐ a. when these models have been learned with different loss functions
- ☑ b. when all models' predictions are mutually independent
- ☑ c. when these models have been learned with different features
- ☐ d. when all models' predictions are highly correlated

Let us assume that a neural network with F inputs is designed to address a multicalss (C>2) classification problem: how many neurons and which type of nonlinearity shall be employed in the output layer ?

Veuillez choisir au moins une réponse :

- ☐ a. F neurons with softmax activation function
- ☐ b. C neurons with sigmoid activation functions
- ☐ c. 1 neuron with sigmoid activation function
- ☑ d. C neurons with softmax activation function

When shall/could the ReLU nonlinearity be employed and where in designing a neural network ?

Veuillez choisir au moins une réponse :

- ☐ a. In the output layer, for classification problems but only if the problem is binary (C=2)
- ☐ b. In the output layer for classification problems
- ☑ c. In hidden layers regardless wether it is a regression or classification problem
- ☑ d. In the output layer, for unbounded regression problems with positive image

Having two imbalanced classes, what should I do ?

Veuillez choisir au moins une réponse :

- ☑ a. change the local loss function and take into account the class in the cost of a wrong prediction
- ☑ b. keep the same loss as usual and apply my preferred algorithm to a modified training dataset that oversamples the rare class
- ☐ c. keep the loss as usual and apply my preferred algorithm to a modified dataset that undersamples the frequent class

In which case the use of a multilayer (i.e., with at least one hidden layer in addition to the output layer) neural network in a classification problem is justified ?

Veuillez choisir au moins une réponse :

- ☐ a. When a classification problem consists in separating the input data in at least C=2 classes
- ☐ b. When a single layer neural network has been verified to overfit on the training data
- ☐ c. When a single layer neural network has been verified not to overfit on the training data
- ☑ d. When the input data cannot be separated by a linear classifier

In which case(s) shall be a regularization term added to the cost function minimized during the training of a neural network ?

Veuillez choisir au moins une réponse :

- ☑ a. When the network tends to overfit over the training data as witnessed by a proper validation set
- ☐ b. None of the other answers are correct
- ☐ c. When the network does not overfit on the training data but still the training error is higher than the validation error
- ☐ d. When the backpropagated error gradients are >>1 and the network fails to converge

Which of these statements on neural networks are true ?

Veuillez choisir au moins une réponse :

- ☒ a. A multilayer perceptron with one hidden layer can approximate any function
- ☑ b. Adding neurons on any layer improves the capacity on the neural network
- ☐ c. 2 perceptrons are enough to learn the XOR function

What are the positive definite symmetric (PSD) kernels among these functions ?

Veuillez choisir au moins une réponse :

- ☐ a. k(x,x')=cos(x-x') is PSD kernel
- ☑ b. k(x,x')=exp(-gamma D(x,x')^2) where D is a distance and gamma > 0 is a PSD kernel
- ☑ c. k(x,x')= [sin(x-y)]^2

Let us assume that a neural network with F inputs is designed to address a binary (C=2) classification problem: how many neurons and which type of nonlinearity shall be employed in the output layer ?

Veuillez choisir au moins une réponse :

- ☐ a. F neurons with softmax activation function
- ☐ b. 2 neurons with softmax activation function
- ☐ c. 2 neurons with sigmoid activation functions
- ☑ d. 1 neuron with sigmoid activation function

How to control the complexity of a decision tree ?

Veuillez choisir au moins une réponse :

- ☐ a. by imposing a number of nodes equal to the training set size
- ☑ b. by giving a lower bound on the size of a leaf (number of training data fallen into this leaf)
- ☑ c. by imposing a maximal depth

Which among those statements about the perceptron algorithm are correct ?

Veuillez choisir au moins une réponse :

- ☑ a. The perceptron algorithm can separate data with a margin.
- ☑ b. In the perceptron algorithm, an update is made only if the perceptron makes a mistake.
- ☑ c. A perceptron is guaranteed to perfectly learn a linearly separable function within a finite number of steps.

Random Forests defined by Breiman are :

Veuillez choisir au moins une réponse :

- ☐ a. a linear combination of decision trees that have to be learned sequentially
- ☑ b. a linear combination of randomized decision trees, each of them trained on a different bootstrap sample
- ☐ c. a random linear combination of decision trees, each of them trained on a different bootstrap sample
- ☑ d. apart trees, random forests can be applied on any classifiers

Which among those statements about the use of minibatches of B (B>1) training data is correct ?

Veuillez choisir au moins une réponse :

- ☑ a. The use of minibatches gives more reliable error gradients
- ☑ b. The use of minibatches usually improves the performance of a trained network because it offers more optimization chances for a given train set and number of training epochs
- ☐ c. The use of minibatches is incompatible with the use of ReLu nonlinearities
- ☑ d. The use of minibatches eliminates or at least reduces the problem with the vanishing gradient

About kernels :

Veuillez choisir au moins une réponse :

- ☐ a. the kernel trick is a computational heuristic to compute faster a nonlinear classifier
- ☐ b. given a kernel there is always a unique pair of (feature map, feature space).
- ☑ c. the kernel trick is the approach that consists in embedding data into a feature space and work in this space without paying the price to compute the inner products

Which of these are valid strategies for choosing the value of the learning rate ?

Veuillez choisir au moins une réponse :

- ☑ a. Choose a initial learning rate and adjust it depending on the inverse of the magnitude of the gradient
- ☐ b. Choose a initial learning rate and decrease it when the validation metric goes up
- ☐ c. Choose a initial learning rate and increase it periodically