

⚠ Students have either already taken or started taking this quiz, so be careful about editing it. If you change any quiz questions in a significant way, you may want to consider regrading students who took the old version of the quiz.

Points 6 ✔ Published

Details

Questions

 Show Question Details

Question

1 pts

In practice, what is the most accurate description for activation functions (such as Sigmoid, Sum, Tanh, ReLU) used in neural networks?

- They must be differentiable.
- They can be non-differentiable, but only for a small amount of points.
- They can be any continuous functions.
- They must be non-linear to be learnable.

Answer

Question

1 pts

Given a neural network with N input nodes, no hidden layers, one output node, with entropy loss and sigmoid activation functions, which of the following algorithms (with the proper hyper-parameters and initialization) can be used to find the global optimum?

- Stochastic Gradient Descent
- Batch Gradient Descent
- Mini-Batch Gradient Descent
- All of the above

Answer

Question

1 pts

Assume that for a group of people, the probability for individuals to get Covid is 1%. Comparatively, a fever is fairly common, and the probability for individuals to get a fever is 10%.

If 90% of the covid patients develop fevers, what is the probability for an individual to be a covid patient if this individual has a fever?

Enter your answer as an integer that represents the percentage of this probability.

Answers 9 (with margin: 0)

Question

1 pts

An old factory that has a defect rate of 50% (50% of the products have some defects) uses a quality assurance check to filter out products with defects. The check has a 90% chance to identify defects in a product if the product is defective. For the well-manufactured products, the check has a 10% chance to falsely identify a defect in a product and filter it out.

If the check filters a product out, what is the probability that it has defects?

Enter your answer as an integer that represents the percentage of this probability.

Answers 90 (with margin: 0)

Question

1 pts

What is the best description of the back-propagation process on the max-pooling operation?



It uses a max() operation which is differentiable, so the back-propagation process is the same as other functions.



It uses a softmax() operation which produces a differentiable estimation of the max value.



Despite being non-differentiable, it can be propagated since it can remember the index that has the max value and only update the weights related to the selected value during back propagation.

Answer



It remembers the index that has the max value and update the other indices in the pooling layer with negative gradients.

Question

1 pts

You want to train a neural network to predict the next 30 daily prices using the previous 30 daily prices as inputs. Which model selection and explanation make the most sense?



A fully connected deep feed-forward network because it considers all input prices in the hidden layers to make the best decision.



A single one-directional RNN because it considers the order of the prices, and the output length is the same as the input length.



A bidirectional RNN because the prediction benefits from future labels.



A one-directional encoder-decoder architecture because it can generate a sequence of future prices based on all input historical prices.

Answer

[+ New Question](#)[+ New Question Group](#)[Find Questions](#) **Notify users this quiz has changed**[Cancel](#)[Save](#)