DM581, AI512: Introduction to Machine Learning - Re-exam February 2024

The exam contains 11 questions. All questions ask for an evaluation of five statements with (yes/no) or (true/false) answers. A true answer adds +1 points and a false answer adds -1 points to your exam score. Hence, skipping to evaluate a statement is very likely to be more advantageous for you than making a random guess. Your final score is calculated as (Your point sum)/(11*5)*100. The exam duration is four hours.

$P(\cdot)$ denotes a probability measure.

No assumptions may be made additional to those specified in the question text.

```
1: for epoch in range(100):
2:     for inputs_batch, outputs_batch in train_dl:
3:         predictions_batch = model(inputs_batch.view(-1, 784))
4:         loss = loss_function(predictions_batch, outputs_batch)
5:         loss.backward()
6:         optimizer.step()
7:         optimizer.zero_grad()
```

Evaluate the following statements related to the Python source code given above which uses the PyTorch library.

Select the correct answers

	True	False
The code performs gradient-descent updates 100 times.	○	○
The code will give a dimensionality mismatch error regardless of the minibatch size because the "inputs_batch" tensor is transformed into an invalid size by the "view(-1, 784)" operation.	○	○
The code implements the training stage of a supervised learning task.	○	○
Line 6 updates the model parameters.	○	○
Line 5 computes the gradients of the loss function for the drawn minibatch.	○	○

```
1: model = nn.Sequential(
2:     nn.Linear(10, 100),
3:     nn.ReLU(),
4:     nn.Linear(100, 250),
5:     nn.ReLU(),
6:     nn.Linear(250, 125),
7:     nn.ReLU(),
8:     nn.Linear(125, 2)
9: )
```

Evaluate the following statements about the PyTorch source code given above and the neural network architecture it builds.

Select the correct answers

	True	False
The neural network uses the hyperbolic tangent activation function in all layers.	○	○
Line 4 should be replaced by nn.Linear(250, 100) for the source code to run without a dimensionality mismatch error.	○	○
The code will run without a dimensionality mismatch error only on minibatches of 10 observations.	○	○
The "forward" function of the created neural network object returns a 125-dimensional output vector.	○	○
The source code implements a neural network architecture for 10-dimensional input observations.	○	○

You are given a k-Nearest-Neighbor (kNN) classifier evaluated on a particular training and test split of a particular data set for a particular choice of k. Evaluate the following statements related to this classifier.

Select the correct answers

	True	False
The kNN classifier's estimator variance gets smaller as the training set size gets smaller.	○	○
The kNN classifier's test accuracy is guaranteed to increase by reducing "k"	○	○
The kNN classifier's prediction error on the training set may be larger than its prediction error on the test set in some data sets.	○	○
The kNN classifier's underfitting risk is mitigated by increasing training data size.	○	○
The kNN classifier's overfitting risk is mitigated by increasing "k".	○	○

Consider a discrete random variable X that follows the probability distribution below:

$$P(X = 1) = 0.2$$
$$P(X = 2) = 0.5$$
$$P(X = 3) = 0.3$$

Evaluate the following statements regarding this random variable.

Select the correct answers

	True	False
The variance of the random variable X is smaller than 0.	○	○
The mean of the random variable X is greater than 2.	○	○
$P(X < 3) = 0.7$	○	○
$P(X < 4) < 1$	○	○
$P(X \text{ is odd}) = P(X \text{ is even})$	○	○

Two independent random variables X_1 and X_2 are known to have the following properties:

$$\mathbb{E}[X_1] = \mu_1$$
$$\mathbb{E}[X_2] = \mu_2$$
$$\mathrm{Var}[X_1] = \sigma_1^2$$
$$\mathrm{Var}[X_2] = \sigma_2^2$$

where $\mathbb{E}[\cdot]$denotes the expectation and $\mathrm{Var}[\cdot]$denotes the variance of a random variable.

Evaluate the following statements about these two random variables.

Select the correct answers

	True	False
$\mathbb{E}[aX_1 + bX_2] = (a+b)(\mathbb{E}[X_1] + \mathbb{E}[X_2])$	○	○
$\mathrm{Var}[c(X_1 + X_2)] = c^2\mathrm{Var}[X_1] + c^2\mathrm{Var}[X_2]$	○	○
$\mathbb{E}[X_1^2 + X_2^2] = \mu_1^2 + \mu_2^2 + \sigma_1^2 + \sigma_2^2$	○	○
$\mathrm{Var}[X_1 \cdot X_2] = \mathrm{Var}[X_1] \cdot \mathrm{Var}[X_2]$	○	○
$\mathrm{Var}[X_2^2] = \mathrm{Var}[X_2]^2$	○	○

For the three events A, B, and C we know that

- A and C are independent
- B and C are independent
- A and B are disjoint
- $P(A \cup B) = x$
- $P(B \cup C) = y$
- $P(A \cup B \cup C) = z$

  for $x, y, z \in [0, 1]$.

Evaluate the statements given below about these three events.

Select the correct answers

	True	False
$P(B) = x + y - z$	○	○
$P(A) = \frac{z-x}{1-(x+y-z)}$	○	○
If $x = y$ then $P(A) = P(C)$.	○	○
$z - x - y + 1 \geq 0$	○	○
$y > z$	○	○

A fair coin is tossed four times. The event that one side of the coin shows upwards when the coin is at rest is denoted as "heads" and the other side as "tails". The coin tosses are independent and identically distributed events. In other words, the outcomes of the previous tosses do not have any effect on the outcome of the next one. Evaluate the following statements regarding this coin toss experiment.

Select the correct answers

	True	False
Probability of at least two heads is equal to the probability of at least two tails.	○	○
The expected value of the number of heads is 1/2.	○	○
Probability of four tails is 1/16.	○	○
Probability of at least one heads is less than the probability of at least two heads.	○	○
"The number of heads being odd" and "getting two or fewer heads" are independent events.	○	○

Evaluate the following statements about the Q-learning algorithm.

Select the correct answers

	True	False
Q-learning is an unsupervised learning algorithm.	○	○
Q-learning assumes that the reward of a state transition resulting from a taken action is observed.	○	○
The output of the Q-learning algorithm run on an episodic environment is a function that maps a state-action pair to an estimate of the sum of discounted rewards until the end of the episode.	○	○
Q-learning assumes that the exact state transition distribution is available.	○	○
Q-learning iterations are guaranteed to converge only if the agent's actions do not affect the environment state.	○	○

Consider the confusion matrix below obtained from a classifier evaluated on the test split of a data set comprising data points with discrete labels that can take three possible values: A, B, C. Each of these values correspond to a class.

		Predicted			
		A	**B**	**C**	**Total**
Actual	**A**	20	10	10	40
	B	8	30	7	45
	C	3	2	10	15
	Total	31	42	27	100

Evaluate the below statements regarding this confusion matrix.

Select the correct answers

	True	False
Precision for Class A is smaller than recall for Class A.	○	○
Precision for Class B is smaller than recall for Class B.	○	○
Precision for Class C is smaller than recall for Class C.	○	○
The random classifier that predicts each of the three classes with equal probability has an expected accuracy of 50% on this test split.	○	○
The accuracy of a classifier that always predicts Class B on this test split is higher than the expected accuracy of a random classifier that predicts each of the three classes with equal probability.	○	○

We train the k-means algorithm with k=2 on a data set comprising the following eight data points with two features each:

D={(1,1),(2,2),(3,1),(4,4),(5,4),(5,5)}

using the Euclidean distance to assign data points to clusters and breaking the ties in favor of Cluster 1 (a data point with equal Euclidean distance to Cluster 1 and Cluster 2 is assigned to Cluster 1).

Evaluate the following statements regarding the outcome of the k-means algorithm when it is run with the following two cluster centroid initialization options:

Option 1:

$\mu_1 = (3, 3)$ for Cluster 1,
$\mu_2 = (3, 4)$ for Cluster 2.

Option 2:

$\mu_1 = (2, 2)$ for Cluster 1,
$\mu_2 = (3, 1)$ for Cluster 2.

Select the correct answers

	True	False
When initialized with Option 2, the algorithm converges after the centroids are updated only once.	○	○
When initialized with Option 1, the data points (1,1),(2,2),(3,1) are assigned to Cluster 1 and (4,4),(5,4),(5,5) are assigned to Cluster 2.	○	○
When initializated with Option 1, the algorithm converges after the centroids are updated only once.	○	○
The two initialization options converge to different clusterings results.	○	○
When initialized with Option 2, the data points (1,1),(2,2),(3,1) are assigned to Cluster 1 and (4,4),(5,4),(5,5) are assigned to Cluster 2.	○	○

Consider the experiment of the toss of three coins. It is given that two of the coins are fair and one is unfair. The unfair coin has a Heads probability of 3/4. Let U be a random variable denoting the event that a coin is unfair. The following is also given:

$$P(U = 1) = 1/3$$
$$P(U = 2) = 1/3$$
$$P(U = 3) = 1/3$$

In words, there is no prior knowledge about which of the three coins is unfair. The first two coins are observed to be Heads (H) and the third coin Tails (T). Denote the outcome of this experiment as HHT.

Evaluate the following statements related this coin toss experiment.

Select the correct answers

	True	False
$P(U = 1 \| HTH) < P(U = 1 \| HHT)$	○	○
$P(U = 2 \mid HHT) > P(U = 2)$	○	○
$P(U = 1 \mid HHT) > P(U = 1)$	○	○
$P(U = 3 \mid HHT) > P(U = 3)$	○	○
$P(U = 2 \| HHT)$ would not change if $P(U = 2)$ were equal to 2/3.	○	○