Started on	Sunday, 21 July 2024, 10:43 PM
State	Finished
Completed on	Monday, 22 July 2024, 1:20 PM
Time taken	14 hours 37 mins
Marks	67/67
Grade	**10** out of 10 (**100**%)

Question 1
Correct
Mark 4 out of 4

- A probability space consists of the set of elementary events, a definition which subsets are measurable, and a [probability measure].
- Measurable subsets of a probability space are called [events].
- Single elements of the probability space are [elementary events].
- Functions mapping elementary elements to elements in another measurable space are called [random variables].
- A function mapping events to probabilities is a [probability measure].
- Features and derived features of data points can also be modeled as [random variables].
- An output value in [0,1] of a probability measure is a [probability].

[probability] [probability measure] [elementary events] [random variables] [events]

Your answer is correct.

Correct
Marks for this submission: 4/4.

Question 2
Correct
Mark 4 out of 4

Consider a probability space (Ω, Σ, P). Consider further a feature X with values in the measurable space (S, Σ_S) (i.e., a random variable $X: \Omega \to S$). Also recall that the probability distribution over feature values of X is defined by the push forward $P \circ X^{-1}: \Sigma_S \to [0,1], A \mapsto P(X \in A) = P(X^{-1}(A))$.

For characterizing/specifying the probability distribution, the following functions are helpful:

$P(X = a)$ (discrete case): For discrete features (e.g., nominal and ordinal ones), the function $S \to [0, \infty), a \mapsto P(X = a)$ mapping discrete feature values to their probability is a [probability mass]

$P: \Sigma \to [0,1]$ is called a [probability measure]

$P(X = a)$ (continuous case): For dense features (i.e., a real-valued random variable), the function $a \mapsto P(X = a) = (\frac{d}{da} P(X \leq a))(a)$ is the [probability density function]

$P(X \leq a)$: For dense features (e.g., interval features), the function $a \mapsto P(X \leq a)$ is the feature's [cumulative distribution function]

Your answer is correct.

Correct
Marks for this submission: 4/4.

Question 3

Correct

Mark 4 out of 4

The following functions measure probabilities ...

- for any event / subset of feature values: probability measure
- for single discrete feature values: probability mass
- for intervals of real-valued feature values: cumulative distribution function (CDF)
- for single continuous feature values (as infinitesimal intervals): probability density function (PDF)

Amongst these, the [[3]] is the cumulated (integral of) probability density.

probability measure	cumulative distribution function (CDF)	probability mass	probability density function (PDF)

Your answer is correct.

Correct

Marks for this submission: 4/4.

Question 4

Correct

Mark 9 out of 9

What of the following statements about value ranges holds true?

True	False		
○	◉	Probability masses / densities are always < 1.	What about a deterministic process producing always the same feature value?
◉	○	Probability masses / densities are always ≥ 0.	
○	◉	Probability densities are always ≤ 1.	Consider the equal distribution over [0, 0.5], i.e., the density maps all values in [0,0.5] to a constant, and values outside of the interval to 0. What value must the constant have in order to ensure that the area under this bump (the integral over all densities) sums up to 1?
◉	○	Cumulative probabilities of intervals are always ≤ 1.	Remember that $a < b \Rightarrow (-\infty, a] \subset (-\infty, b] \Rightarrow \{X \leq a\} \subset \{X \leq b\}$ and that $A \subset B \rightarrow P(A) \leq P(B)$; what does this mean if $P(\Omega) = P(X \leq \infty) = 1$?
◉	○	Cumulative probabilities of intervals are always ≥ 0.	
○	◉	Probability masses / densities are always > 0.	What about the unfair dice that only produces six eyes?
◉	○	Probabilities are always in [0,1].	This is given by the definition.
◉	○	Probability masses are always ≤ 1.	Can the sum of all masses be 1 if all values are ≥ 0 and one value is > 1?
○	◉	Cumulative probabilities of intervals are always > 0.	

Correct

Marks for this submission: 9/9.

Question 5

Correct

Mark 4 out of 4

ERRATA: The answer to $P_X = P(X,y) \forall y \Rightarrow X,Y \text{independent}$ should be TRUE. This is wrongly marked as false here (and will not be corrected, as the quiz has been submitted with this faulty options already). Find a deduction of this fact in the script following the definition of marginalization.

We have seen that one can consider multivariate distributions, i.e., the distribution of several random variables (alias feature values) simultaneously. Both marginalization and conditional distribution pose ways to extract from a given multivariate distribution information about the distribution of a single feature value. In the 2D example, we want to reduce the multivariate distribution P of features X, Y on $\mathbb{R}^2$ to a distribution of features X:

- Conditional distribution density at $Y = y$: $a \mapsto P(X = a \mid Y = y) = \frac{P(X=a, Y=y)}{P(Y=y)}$
- Marginal distribution density: $a \mapsto P_X(X = a) = \int_y P(X = a, Y = y)dy = \int_y P(X = a \mid Y = y)P(Y = y)dy$

The main difference is, that the conditional distribution is the distribution of X on those points that fulfill $Y = y$ (the vertical cut of the scatter plot where $Y = y$); and for marginalization all such "cuts" are averaged. Some typical cases are:

- The special case that X, Y are independent means that the density $P(X = \cdot \mid Y = y)$ is independent of the value of Y – in formulas this is exactly the definition of independence: $P(X = \cdot \mid Y) = P(X)$ (and the other way round).
- A non-zero covariance captures exactly the fact that X depends on Y to some extend.

Now consider a multivariate distribution of two variables X, Y. Which of the following statements are true?

True	False	
(x)	( )	If the variables are independent, their covariance is 0.
( )	(x)	If the variables have a covariance of 0, they are independent.
(x)	( )	If the variables are independent, the marginal distribution P_X is for any value y of Y equal to the conditional distribution $P(X \mid Y = y)$.
( )	(x)	If the marginal distribution P_X is equal to $P(X \mid Y = y)$ for any value y of Y, the variables are independent.

Correct

Marks for this submission: 4/4.

Question 6

Correct

Mark 11 out of 11

Some very kind person provided you with gummi bears, which have colors red and yellow. You would like to describe the stochastic process of drawing from this set of differently colored gummi bears. Let's consider as event the drawing of a color.

Simple case: Two colors, single draw, uniform distribution

Assume there are only two colors, i.e., the probability space is $\Omega = \{\text{red}, \text{yellow}\}$ with valid events $\Sigma = \{\{\}, \{\text{red}\}, \{\text{yellow}\}, \{\text{red}, \text{yellow}\}\}$.

If the colors are equally distributed, the probability of drawing a red bear is: 0.5

More colors, single draw, uniform distribution

Reconsider the modeling: There are also green and orange gummi bears, and green is your favorite color.

Still assuming equal distribution, the probability of not drawing your favorite: $P(\{\text{red}, \text{yellow}, \text{orange}\}) =$ 0.75

Correct

Marks for this submission: 11/11.

Question 7

Correct

Mark 2 out of 2

(continued)

Independent random variables (equal distribution)

Go for another gummi bear: We now consider two independent draws (assume there is an endless supply of gummi bears in all colors for you).
Our probability space now is Ω^2. Independence here means that for any $A, B \subset \Omega$: $P(A \times \Omega \cup \Omega \times B) = P(A \times \Omega) \cdot P(\Omega \times B)$. Speaking in terms of random variables: The projections $X_1 : \Omega^2 \to \Omega, (x_1, x_2) \mapsto x_1$ and $X_2 : (x_1, x_2) \mapsto x_2$ to the first resp. second draw result are independent.

Given that,

- the probability of drawing two bears of your favorite color is $P(X_1 = \text{green}, X_2 = \text{green}) =$ 0.0625
- the probability of drawing two bears both not of your favorite color is $P(X_1 \neq \text{green}, X_2 \neq \text{green}) =$ 0.5625

Correct

Marks for this submission: 2/2.

Question 8

Correct

Mark 3 out of 3

(continued)

Independent random variables (non-uniform distribution)

You are not the only one loving green: The production lines produce double the amount of green bears compared to any other color (probability of drawing green in a single draw is 0.4).

- The probability of drawing exactly two bears of your favorite color is: 0.16
- The probability of getting at least one green gummi bear when drawing two bears is
  $P(\{X_1 = \text{green}\} \cup \{X_2 = \text{green}\}) = P(\{\text{green}\} \times \Omega \cup \Omega \times \{\text{green}\}) =$ 0.64

  Tip: You can here use the identity $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ or $P(\neg A) = 1 - P(A)$.
- The probability of getting exactly one green gummi bear when drawing two is $P((\{X_1 = \text{green}\} \cup \{X_2 = \text{green}\}) \setminus \{(\text{green}, \text{green})\}) =$ 0.48

  Tip: You have calculated $P(\{\text{at least one green}\})$ and $P(\{\text{exactly two green}\})$. You can now use the identity $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$ and the fact that $\{\text{exactly one green}\} \cup \{\text{exactly two green}\} = \{\text{at least one green}\}$ and $\{\text{exactly one green}\} \cap \{\text{exactly two green}\} = \emptyset$.
  Alternatively, you can reformulate this with conditional probabilities
  $P(X_1 \neq \text{green}, X_2 = \text{green}) + P(X_1 = \text{green}, X_2 \neq \text{green}) = P(X_2 = \text{green} \mid X_1 \neq \text{green})P(X_1 \neq \text{green}) + P(X_2 \neq \text{green} \mid X_1 = \text{green})P(X_1 = \text{green}$
  and use the independence of the variables to determine the respective values.

Correct

Marks for this submission: 3/3.

Question 9

Correct

Mark 4 out of 4

(continued)

Dependent random variables

Your previously (as good as) exhaustless gummi bear stock grew quite small: Only 5 bears left, two green and one of each other (i.e., probabilities of first draw stay the same). This makes the second draw very dependent on the first, as in the first you still have five bears, in the second only four left. For example, drawing a red gummi bear in the first draw remains $P(X_1 = \text{red}) = \frac{1}{5} = 0.2$; but given that you drew a green one in the first round, the probability of drawing the red one in the second round is $P(X_2 = \text{red} \mid X_1 \neq \text{red}) = \frac{1}{4}$.

- The probability of drawing both green gummi bears now is 0.1

  Tip: You can here use that $P(A \mid B) = \frac{P(A,B)}{P(B)}$, i.e. $P(\{X_2 = \text{green}\} \cap \{X_1 = \text{green}\}) = P(X_2 = \text{green} \mid X_1 = \text{green}) \cdot P(X_1 = \text{green})$. The conditional probability can be derived as shown above.
- The probability of drawing a green gummi bear only in the first round is:

  $P((X_1 = \text{green}), (X_2 \neq \text{green})) = P(X_2 \neq \text{green} \mid X_1 = \text{green}) P(X_1 = \text{green}) =$ 0.3
- The probability of drawing a green gummi bear only in the second round is:

  $P((X_1 = \text{green}), (X_2 \neq \text{green})) = P(X_2 = \text{green} \mid X_1 \neq \text{green}) P(X_1 \neq \text{green}) =$ 0.3
- The probability of drawing exactly one green gummi bear is $P(\{\text{only first round}\}) + P(\text{only second round}\}) =$ 0.6

Correct

Marks for this submission: 4/4.

Question 10

Correct

Mark 4 out of 4

Let's revisit the example from the lecture: The testing of a binary classifier can be modeled using two random variables, one being the the predictor Pred and one the ground truth labels Pos.

We saw how to calculate the positive predictive value $P(\text{Pos} \mid \text{Pred})$ from recall, false positive rate, and the probability distribution of Pred in the form of $P(\text{Pred}), P(\neg\text{Pred})$. In the case of cancer prediction, the positive predictive value tells, with what probability one has cancer when this was predicted by the predictor.

Let's this time assume that the predictor has a very good recall of 0.99 (nearly all cases of cancer are revealed), sacrificing the false positive rate which is at 0.5 (half of the alarms are false alarms). To get some insight into the influence of the balancing of the ground truth classes, calculate the positive predictive value (PPV) for the following cases:

Positive Predictive Values

	P(Pos)	1-P(Pos)	PPV (+-0.01)
Balanced	0.5	0.5	0.66
Slightly imbalanced	0.4	0.6	0.57
Imbalanced	0.1	0.9	0.18
Strongly imbalanced	0.01	0.99	0.02

Correct

Marks for this submission: 4/4.

Question 11

Correct

Mark 4 out of 4

Which of the following statements is true?

True	False		
◉	○	The variance and standard deviation of a feature both tells how spread its values are.	
◉	○	At the mode of a distribution is the (global) maximum of its density.	
○	◉	The mean of feature values is the same as their expected value.	This only holds for equal distribution.
◉	○	The span requires the minimum and maximum to exist.	
◉	○	The expected value is the weighted average of feature values.	The weights are the probability density / mass.
◉	○	The variance is the square of the standard deviation.	
○	◉	The span requires the variance to exist.	
○	◉	The variance is different name of the standard deviation.	

Correct

Marks for this submission: 4/4.

Question 12

Correct

Mark 4 out of 4

Let $M = \left(\text{Cov}(X_{i}, X_j)\right)_{i,j}$ be the covariance matrix of a family of random variables $X_1, \dots, X_n$, containing the covariances of pairs of the variables. Which of the following statements are correct?

Select one or more:

- [ ] a. The respective correlation matrix is the same as the covariance matrix up to a factor.
- [x] b. The diagonal entries $M_{i,i}$ are the variances of the random variables X_i.
- [x] c. This matrix is symmetric.
- [ ] d. All entries in this matrix must be positive.

Your answer is correct.

Correct

Marks for this submission: 4/4.

Question 13

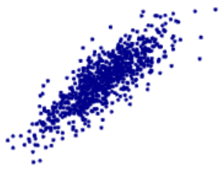
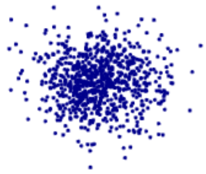
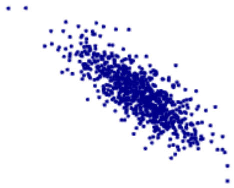
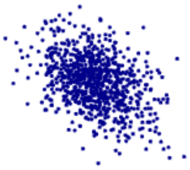
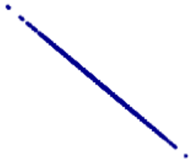
Correct

Mark 4 out of 4

Let's get a better gut feeling of correlation between two variables. For this, we consider several exemplary bivariate distributions (i.e., ones with two random variables \(X, Y\)). For each, a sample is provided in the following and plotted as scatter plot for visualization.

Assign the right correlation values between \(X, Y \) to the different datasets of feature values.

Distributions with Different Correlation

	Corr		Corr		Corr		Corr
	-> 0.8		-> 0		-> 0		-> 0
	-> -0.8		-> -0.4		-> -1		-> 0

0.8 0 -0.4 -0.8 -1

Your answer is correct.

Correct

Marks for this submission: 4/4.

Question 14

Correct

Mark 6 out of 6

Remember that usually one has only access to randomly sampled samples (datasets of observations) of finite size in order to estimate statistics of the underlying population. Since these are only estimations, it is important to keep in mind that these estimations can be erroneous – and the smaller the sample size, the lower the confidence that the estimate from the sample are well matching the actual population statistics.
Which of the following statements are true?

True	False		
○	◉	When taking finite subsamples of a population, the respective sampling distribution is the distribution of sampled features.	
◉	○	When taking finite subsamples of a population, the respective sampling distribution is the distribution of estimations for a population statistics.	
◉	○	The type of the sampling distribution is always a normal distribution.	This is because the sampling process is random.
○	◉	The type of the sampling distribution depends on the distribution of the underlying population.	
◉	○	The variance of the sampling distribution depends on the variance of the population.	
○	◉	The variance of the sampling distribution is independent of the sample size.	
◉	○	To improve standard deviation by a factor of \(k \), one requires \(k^2 \) times more samples.	
◉	○	The expected value of the mean of samples is the mean of the population (i.e., the sampling mean is an unbiased estimator for the population mean).	
○	◉	The standard deviation of samples is an unbiased estimator for the population standard deviation.	
◉	○	The corrected variance of samples \(\frac{n}{n-1}\text{Var}(X) \) is an unbiased estimator of the population variance.	
○	◉	A confidence interval with confidence level \(\alpha \) means: The true population mean lies within the confidence interval around the mean of a sample with probability \(\alpha \).	
◉	○	A confidence interval with confidence level \(\alpha \) means: The true population mean lies within the confidence interval around the mean of a sample with probability \(1-\alpha \).	

Correct

Marks for this submission: 6/6.

◄ 01. Quiz - Deep Neural Networks and Gradient Descent

Jump to...

03. Quiz - Hypothesis Testing ►