COURSE WARM-UP INDIVIDUAL EVALUATION

1. What is a probability density function? Give a concrete example, try to use formulas. (Correct definition 1p, every correct example (up to 3) +1p) – max 4p

    https://en.wikipedia.org/wiki/Probability_density_function

    BLUF: A PDF is a function which calculates the likelihood that a random variable's value will fall within a range of values. For example, this could compute the probability that tomorrow's high temperature might fall within a range(15C °, 20C °)

    Intuitively, one can think of $f_X(x)\,dx$ as being the probability of X falling within the infinitesimal interval $[x, x + dx]$

    From Integral calculus, the area under a curve over a given interval can be calculated by:

    $$\Pr[a \le X \le b] = \int_a^b f_X(x)\,dx.$$

2. What is a probability mass function? Give a concrete example, try to use formulas. (correct definition + example 1+1 p each) – max 2p

    https://en.wikipedia.org/wiki/Probability_mass_function

    BLUF: A PMF is a function which computes the probability that a discrete random variable is equal to an exact value rather than the probability that it falls within a range as in a PDF. An example of this would be to predict whether or not tomorrow's high temperature will be exactly 17 C °.

    The formal proof is thus: $f_X(x) = \Pr(X = x) = P(\{s \in S : X(s) = x\})$

    So, breaking down what this means. This basically says that, to calculate this probability, it helps to think in terms of probability as mass a.k.a rather than probability being equal to the area of a line under a curve, which is 0, it's more important to think that the probability of this exact value is related to the number of occurrences of this exact phenomena on this exact day over a given range of years. Taking that value and dividing it over the total number of days of that range.

3. What is statistical inference and how does it relate to machine learning? Make use of the term 'generalization error' in your explanation. If you don't know what it means, explain in another way, but try to be as precise as possible. (2p for correct answer, 3p if generalization error is used correctly).

    https://stats.stackexchange.com/questions/130867/inference-vs-estimation

BLUF: Statistical inference is the act or process of deriving logical conclusions from premises known/assumed to be true through the use of mathematics. A generalization error or out of sample error is a measure of how accurately a given machine learning algorithm is able to predict outcome values for previously unseen data. These two ideas come together during supervised machine learning which refers to a task of a learning function to map inputs to specified output pairings based on an example provided by training data.The previous two questions are both examples of statistical inference and its application to machine learning/data science as well as the application of supervised learning whereby historical weather analysis is used as the basis of prediction.

4. Please describe in detail what are the assumptions for linear regression. (1p for every correct assumption) (max 6p)

     Week 6: Assu
       Regression

     1. The distribution of residuals is normal (at each value of the dependent variable).

     2. The variance of the residuals for every set of values for the dependent variable is equal (homoscedasticity).
              - The opposite is heteroscedasticity

     3. The error term is additive

     4. At every value of the dependent variable, the expected (mean) value   ofthe residuals is zero.
              - No non-linear relationships

     5. Expected correlation between residuals, for any two cases, is 0 (The   independence a ssumption or lack of autocorrelation.

     6. All independent variables are uncorrelated with the error term.

     7. No independent variables are a perfect linear function of the other   independent variables (no perfect multicollinearity).

     8. The mean of the error term is zero.

5. What is a cost function in machine learning and optimization? Give examples using formulas. (1p for correct definition, +1p for every correct example for cost function) – max 5p

     The cost function measures how well our algorithm maps the target estimate or how well our algorithm performs optimization problems.

```
# Finding the minimum of the function
y = x^2 - 4x + 2
dy/dx = 0 = 2*x - 4 # This is our cost function
x = 2
```

$y = x^{**}2 - 4^*x + 2$
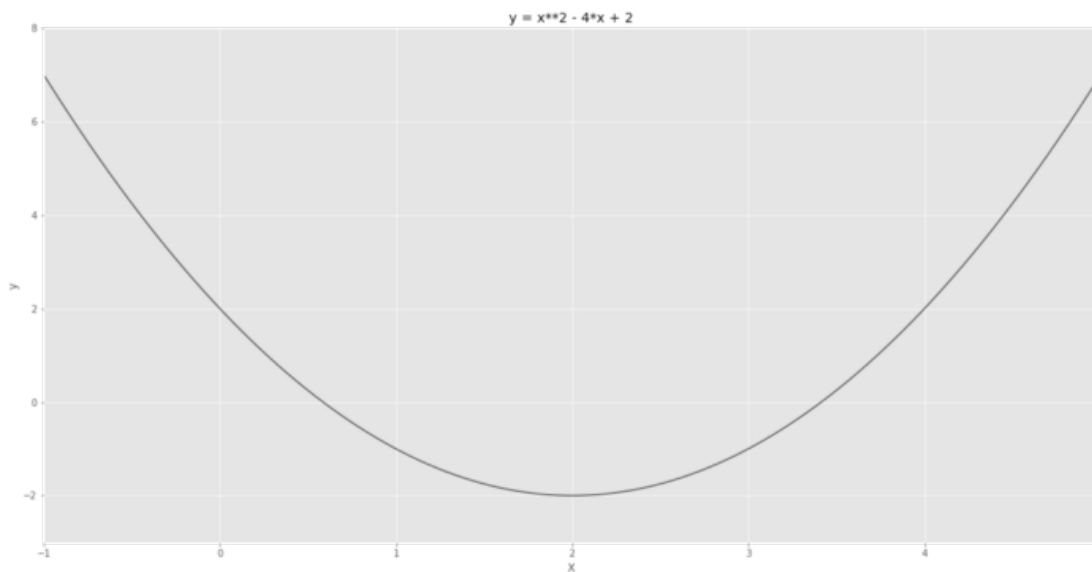
So, essentially, the cost function is really the first derivative of the original function which calculates the slope of the tangent line at a give point or even the average slope over a range of value pairs. Thus, with optimization, we're looking for the point where the slope either becomes zero or at least gets infinitesimally close to zero if the slope shoots to abs(∞).

6. What is bias-variance trade-off? (3p for a good answer)

   https://machinelearningmastery.com/gentle-introduction-to-the-bias-variance-trade-off-in-machine-learning/

   Bias-Variance_Trade-off is the balance between bias built into an algorithm and the variance in its output. The ultimate goal of supervised machine learning is to achieve the lowest possible bias and variance of the output.

7. What is an activation function and what are its common types? (1p for correct definition, +1p for every correct type + 0.5p for every correct formula) – max 5.5p

   https://medium.com/@abhigoku10/activation-functions-and-its-types-in-artifical-neural-network-14511f3080a8

   An activation is a function within a neural network which decides if a neuron should be activated or not. Without these functions, the neural network would simply serve as a linear

regression model incapable of adjusting weighting values and thus incapable of learning from non-linear data models/inputs.

Sigmoid Activation function: It is a activation function of form f(x) = 1 / 1 + exp(-x).

Hyperbolic Tangent function- Tanh : It's mathamatical formula is f(x) = 1—exp(-2x) / 1 + exp(-2x).

ReLu- Rectified Linear units : R(x) = max(0,x) i.e if x < 0 , R(x) = 0 and if x >= 0 , R(x) = x.

8. What is and when should one use unsupervised learning? (1p for correct definition, +1p for every new example type) – max 5p

    https://www.freecodecamp.org/news/when-to-use-different-machine-learning-algorithms-a-simple-guide-ba615b19fb3b/

    Unsupervised learning is where you allow the machine learning algorithm to start learning and outputting a result without any explicit human processing of the data beforehand.

    https://towardsdatascience.com/supervised-vs-unsupervised-learning-14f68e32ea8d

    It is appropriate to use unsupervised learning under two conditions. 1) is when we wish to learn the structure of our data without having explicit labels. 2) when we wish to test our data model's ability to map the inputs to the expected outputs.

9. According to your understanding, what is the main 2 differences between frequentist and Bayesian statistics? (2p for a good answer)

    https://www.probabilisticworld.com/frequentist-bayesian-approaches-inferential-statistics/

    In short, according to the frequentist definition of probability, only repeatable random events (like the result of flipping a coin) have probabilities. These probabilities are equal to the long-term frequency of occurrence of the events in question. Frequentists don't attach probabilities to hypotheses or to any fixed but unknown values in general.

    As a Bayesian, you can use probabilities to represent the uncertainty in any event or hypothesis. Here, it's perfectly acceptable to assign probabilities to non-repeatable events, such as Hillary Clinton winning the US presidential race in 2016. Orthodox frequentists would claim that such probabilities don't make sense because the event is not repeatable. That is, you can't run the election cycle an infinite number of times and calculate the proportion of them that Hillary Clinton won.

10. If you train a machine learning model to predict some phenomenon, then the probability that a machine learning model outputs is the actual probability of the phenomenon's occurrence? True or False? (1p if you get it right, 2p if you can argue why this)

The simple answer to this question is no. The reason for this is that it depends on why your algorithm processed/analyzed the data set. It is likely that the probability output is related to the probability that the output falls within a set tolerance.

11. What is precision and how is it computed? (1p for correct idea, +1p for formula) – max 2p

Precision is how close a measurement comes to another measurement. Precision is determined by a statistical method called a standard deviation. Standard deviation is how much, on average, measurements differ from each other. High standard deviations indicate low precision, low standard deviations indicate high precision.

% error = (accepted - experimental) / accepted *100%

12. What is recall and how is it computed? (1p for correct idea, +1p for formula) – max 2p

Recall(also known as sensitivity) is the fraction of relevant instances that have been retrieved over the total amount of relevant instances

13. What is the relation between sensitivity, specificity and precision and recall? – max 4p

Precision is how close the test outputs are to the expected outputs from the input data. Sensitivity is computed similarly while recall compares the outputs the ALL historical outputs.

14. Please describe the difference between an object and class in object-oriented programming. (1p for correct answer, +1p if you can give a specific example) – max 2p

An object is an instance of a class. Previously defined in the programming language. Think of the class as defining what the object will know and do.

15. Predict the correct output of the following code: >>>a=np.arange(6) >>>print(a) (1p)

[0 1 2 3 4 5]

16. Give the formula for Pearson's correlation coefficient (1.5 p)

$$r = \frac{N\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{[N\Sigma x^2 - (\Sigma x)^2][N\Sigma y^2 - (\Sigma y)^2]}}$$

Where:

N	=	number of pairs of scores
Σxy	=	sum of the products of paired scores
Σx	=	sum of x scores
Σy	=	sum of y scores
Σx^2	=	sum of squared x scores
Σy^2	=	sum of squared y scores

17. What is the difference between K-Means and K-Nearest Neighbours? (+1p for correct definition of each and +2p for a good comparison) – max 4p

http://www.devcoons.com/difference-k-means-k-nearest-neighbor-algorithm/

In short, the algorithms are trying to accomplish different goals. K-nearest neighbor is a subset of supervised learning classification (or regression) algorithms (it takes a bunch of labeled points and uses them to learn how to label other points). It is supervised because you are trying to classify a point based on the known classification of other points. In contrast, K-means is a subset of unsupervised learning clustering algorithms(it takes a bunch of unlabeled points and tries to group them into clusters). It is unsupervised because the points have no external classification.

PROGRAMMING

1. (2p) Write a recursive function that expects an integer 'n' from user between 0 and 10. If the value of the integer is less than 5 then it's value has to be increased recursively until it becomes at least 8. In this case, the program has to terminate and print "Course passed". Otherwise, the value of the integer has to be increased and the program must not terminate. At every step before further action, the current value of the integer (that started when user gave the input 'n') must be printed out.
2. (2p) Write a recursive function for implementing the factorial operator. If you don't know what the factorial is, then given any positive integer n, factorial is defined as the product 1x2x…xn.

3. (3p) A non-empty zero-indexed array A consisting of N numbers is given. The array is sorted in non-decreasing order. The absolute distinct count of this array is the number of distinct absolute values among the elements of the array.

For example, consider array A such that:

A[0] = -5 A[1] = -3 A[2] = -1 A[3] = 0 A[4] = 3 A[5] = 6 The absolute distinct count of this array is 5, because there are 5 distinct absolute values among the elements of this array, namely 0, 1, 3, 5 and 6.

Write a function:

def solution(A)

that, given a non-empty zero-indexed array A consisting of N numbers, returns absolute distinct count of array A.

For example, given array A such that:

A[0] = -5 A[1] = -3 A[2] = -1 A[3] = 0 A[4] = 3 A[5] = 6 the function should return 5, as explained above.

Assume that:

N is an integer within the range [1..100,000]; each element of array A is an integer within the range [−2,147,483,648..2,147,483,647]; array A is sorted in non-decreasing order.

Complexity:

expected worst-case time complexity is O(N); expected worst-case space complexity is O(N), beyond input storage (not counting the storage required for input arguments).

HINTs: Which data structure has only unique elements? Can a mathematical operator be used here also?

3. (5p) A zero-indexed array A consisting of N integers is given. A triplet (P, Q, R) is triangular if 0 ≤ P < Q < R < N and:

A[P] + A[Q] > A[R],

A[Q] + A[R] > A[P],

A[R] + A[P] > A[Q].

For example, consider array A such that:

   A[0] = 10   A[1] = 2   A[2] = 5

   A[3] = 1   A[4] = 8   A[5] = 20

Triplet (0, 2, 4) is triangular.

Write a function: def solution(A)

that, given a zero-indexed array A consisting of N integers, returns 1 if there exists a triangular triplet for this array and returns 0 otherwise.

For example, given array A such that:

   A[0] = 10   A[1] = 2   A[2] = 5

   A[3] = 1   A[4] = 8   A[5] = 20

the function should return 1, as explained above. Given array A such that:

   A[0] = 10   A[1] = 50   A[2] = 5

   A[3] = 1

the function should return 0.

Assume that:

N is an integer within the range [0..100,000]; each element of array A is an integer within the range [−2,147,483,648..2,147,483,647].

Complexity:

expected worst-case time complexity is O(N*log(N));

expected worst-case space complexity is O(N), beyond input storage (not counting the storage required for input arguments).

4. (2p)

A zero-indexed array A consisting of N different integers is given. The array contains integers in the range [1..(N + 1)], which means that exactly one element is missing.

Your goal is to find that missing element.

Write a function:

def solution(A)

that, given a zero-indexed array A, returns the value of the missing element.

For example, given array A such that:

A[0] = 2 A[1] = 3 A[2] = 1 A[3] = 5 the function should return 4, as it is the missing element.

Assume that:

N is an integer within the range [0..100,000]; the elements of A are all distinct; each element of array A is an integer within the range [1..(N + 1)]. Complexity:

expected worst-case time complexity is O(N); expected worst-case space complexity is O(1), beyond input storage (not counting the storage required for input arguments).

5. (4p)

A zero-indexed array A consisting of N integers is given. Rotation of the array means that each element is shifted right by one index, and the last element of the array is moved to the first place. For example, the rotation of array A = [3, 8, 9, 7, 6] is [6, 3, 8, 9, 7] (elements are shifted right by one index and 6 is moved to the first place).

The goal is to rotate array A K times; that is, each element of A will be shifted to the right K times.

Write a function:

def solution(A, K)

that, given a zero-indexed array A consisting of N integers and an integer K, returns the array A rotated K times.

For example, given

   A = [3, 8, 9, 7, 6]

   K = 3

the function should return [9, 7, 6, 3, 8]. Three rotations were made:

   [3, 8, 9, 7, 6] -> [6, 3, 8, 9, 7]

   [6, 3, 8, 9, 7] -> [7, 6, 3, 8, 9]

   [7, 6, 3, 8, 9] -> [9, 7, 6, 3, 8]

For another example, given

   A = [0, 0, 0]

   K = 1

the function should return [0, 0, 0]

Given

A = [1, 2, 3, 4]

K = 4

the function should return [1, 2, 3, 4]

Assume that:

N and K are integers within the range [0..100]; each element of array A is an integer within the range [−1,000..1,000]. In your solution, focus on correctness. The performance of your solution will not be the focus of the assessment.

MAX POINTS: 70