## Question 1 — 0.5 pts

Using Machine Learning (ML) to separate salmon from sea bass is an example of what?

- ○ Recession
- ○ Regression
- ○ Assimilation
- ○ Classification

## Question 2 — 0.5 pts

Fitting a curve to some data to use the curve to predict real values (e.g., temperature, probability, etc.) is called what?

- ☐ Classification
- ☐ Regression
- ☐ Regularization
- ☐ Aggression

## Question 3 — 0.5 pts

Model parameters are conventionally denoted by what?

- ○ Theta
- ○ Neta
- ○ Beta
- ○ David Guetta

## Question 4 — 0.5 pts

As bad as bias can be, there's one kind of bias that's absolutely essential to ML, without which learning cannot happen. What's it called?

- ○ Induction Bias
- ○ Induction Motors

- ○ Indiana Jones
- ○ Inductive Bias

## Question 5 — 0.5 pts

How often **during your model development effort** should you open your Test Data (held-out data) to inspect every row of data just to make sure data has remained uncorrupted while living on your hard drive? **Select all that apply.**

- ☐ Never!
- ☐ Never!
- ☐ Never!
- ☐ Never!

## Question 6 — 0.5 pts

Probabilities that make up a complete probability distribution ____.

- ○ could add up to any value.
- ○ must always add up to 1.
- ○ could add up to 0.
- ○ must always add up to 100.

## Question 7 — 0.5 pts

$P(A \cap B) = P(A)\,P(B \mid A)$ is called the ____ for two events.

- ○ Chain Rule
- ○ Backward-chaining Rule
- ○ Chaining
- ○ Forward-chaining Rule

## Question 8 — 0.5 pts

A Gaussian Distribution can be parameterized or defined using two parameters. What are they?

- ○ The Mean and the Median
- ○ The Median and the Standard Deviation
- ○ The Mean and the Mode
- ○ The Mean and the Variance

## Question 9 — 0.5 pts

If A and B are independent events, then $P(A \cap B) = P(A)\,P(B)$.

- ○ Yep
- ○ What?!
- ○ Nope

## Question 10 — 0.5 pts

What is the probability of a model given data called? In other words, what is P(model | data) called?

- ○ Posterior
- ○ Poster
- ○ Prior
- ○ Likelihood

## Question 11

0.5 pts

Bayes' Theorem lets us update our prior belief given new data or observations. True or false?

- ○ True
- ○ False

## Question 12

0.5 pts

What is the probability of data given a model called? In other words, what is P(data | model) called?

- ○ Prior
- ○ Likelihood
- ○ A Posteriori
- ○ Posterior

## Question 13

0.5 pts

Bayesian probability requires that we have a ____ belief.

- ○ blind
- ○ prior
- ○ posterior
- ○ biased

## Question 14 — 0.5 pts

When we compare the posterior probabilities of two models as calculated using Bayes' Theorem, we can do away with one of the components of the Theorem, i.e., we can omit it from the calculations. Which component is it?

- ○ Normalization Constant
- ○ Likelihood
- ○ Posterior
- ○ Prior

## Question 15 — 0.5 pts

Say we're given the Likelihood---i.e., the probability of some observations (data) given a model---for two models. For example, we're given:

P(stomachache | overeating) = 0.2

P(stomachache | stomach_flu) = 0.6

We can immediately conclude stomachache was probably due to stomach flu, and shouldn't need to ask for more information.

- ○ Correct
- ○ Nobody gets a stomachache from overeating.
- ○ No, both overeating and stomach flu are equally likely, as far as we can tell.
- ○ Nope, we should want to find out what the prior probabilities are for overeating and stomach flu. We can only come to a conclusion as described in the question if we can't find such Priors.

## Question 16 — 0.5 pts

**max** and **argmax** are the same thing. True or false?

- ○ True
- ○ False

## Question 17 — 0.5 pts

The naive assumption that the Naive Bayes Classifier makes is that...

- ...the classes are independent of each other given the features.
- ...the classes are independent of the features.
- ...the features are independent of each other given the class.
- ...the features depend on the class.

## Question 18

0.5 pts

Logarithms are monotonic, so the maximum value of the probability density function (PDF) coincides with the maximum value of the log of the same PDF.

- True
- False

## Question 19

0.5 pts

The only difference between the Maximum a Posteriori (MAP) estimate of a model and the Maximum Likelihood Estimate (MLE) of the same model is that...

- ...the MAP estimate includes the Prior.
- ...the MAP estimate includes the Posterior.
- ...the MLE includes the Prior.
- ...the MAP estimate is less than the MLE.

## Question 20 — 0.5 pts

Which of the following is/are okay to say?

A. Likelihood of some data given a model

B. Likelihood of a model given some data

- ○ A
- ○ B
- ○ Both A and B
- ○ None

## Question 21 — 0.5 pts

In a Bayes Net (BN), the Joint Probability Distribution (JPD) over the random variables that make up the nodes of the BN is represented compactly by dividing it up into several Conditional Probability Tables (CPTs), one for each random variable given only its ancestors.

- ○ True
- ○ False

## Question 22 — 0.5 pts

In a Bayes Net, each node is **conditionally** independent of its **non-descendants** given its **parents**.

- ○ True
- ○ False

## Question 23 — 0.5 pts

A Bayes Net is necessarily a tree. True or false?

- ○ True
- ○ False

## Question 24

0.5 pts

What in the ML pipeline is the first thing to understand and understand really, really well? (Select all that apply.)

- [ ] The data
- [ ] The data
- [ ] The data
- [ ] The data

## Question 25

1 pts

A new genetic test can identify susceptibility to anemia in adults. Only 1% of adults suffer from anemia. Of those adults who have anemia, 90% will test positive using the new genetic test, and the others will test negative. Also, 5% of those who do not have the condition will also test positive using the test, while the others will test negative. A person just tested positive for anemia. What is the probability that they have anemia?

- ○ 0.900
- ○ 0.009
- ○ 0.154
- ○ 0.100

## Question 26

0.5 pts

Two Perceptrons that have the exact same structure---two inputs and one output---can perform two different logical operations, e.g., AND and OR, because of a difference in what?

- ○ Their bias
- ○ Their input weights and bias weight
- ○ Their architecture
- ○ Their inputs

## Question 27 — 0.5 pts

The bias input of a Perceptron is always the same. What is that value?

- ○ 1
- ○ 0
- ○ -1

## Question 28 — 0.5 pts

A (single) Perceptron can only classify _____ classifiable data.

- ○ linearly
- ○ logarithmically
- ○ continuously
- ○ exponentially

## Question 29 — 0.5 pts

Which of the following functions CANNOT be represented using a single Perceptron? (Select all that apply.)

- ☐ OR
- ☐ AND
- ☐ XOR
- ☐ NOT

## Question 30 — 0.5 pts

In every node of a neural network, an activation function like a Sigmoid is used because for neural networks, _____ is key.

- ○ sustainability
- ○ differentiability
- ○ linearity
- ○ explainability

## Question 31 — 0.5 pts

In the training phase of a neural network, when the network has seen every example once, i.e., it has seen the entire training set once, it is called one _____ of training.

- ○ series
- ○ timeframe
- ○ epoch
- ○ episode

## Question 32 — 0.5 pts

The labels in the training set for a neural network are typically encoded as _____.

- ○ hot vectors.
- ○ one-shot vectors.
- ○ one-hot vectors.
- ○ two-hot vectors.

## Question 33 — 0.5 pts

Sigmoid, tanh, and Leaky ReLU are examples of what?

- ○ Activated Function
- ○ Activation Function
- ○ Aggregation Function
- ○ Action Function

## Question 34

0.5 pts

For training a neural network, what's one common name for the objective function?

- ○ Lossy Function
- ○ Loss Function
- ○ Subjective Function
- ○ Logarithmic Function

## Question 35

0.5 pts

Early stopping is one way to prevent what?

- ○ Gradient descent
- ○ Hyperfitting
- ○ Underfitting
- ○ Overfitting

## Question 36

0.5 pts

When it comes to image recognition and vision, what type of neural network is typically used?

- ○ Revolutional Neural Network
- ○ Convolved Neural Network
- ○ Convolutional Neural Network
- ○ Convoluted Neural Network

## Question 37

0.5 pts

The grids of values that are "scanned" over an image in a Convolutional Neural Network (aka ConvNet or CNN) are known as feature maps or filters or _____.

- ○ girdles
- ○ kernels
- ○ griddles

- colonels

## Question 38 — 0.5 pts

When a method like Word2Vec is used to encode a word into a vector, that vector representation is called what?

- A one-hot vector
- A word embodiment
- A tokenization
- A word embedding

## Question 39 — 0.5 pts

Word embeddings produced by Word2Vec can be added and subtracted to produce results that correspond to other word embeddings in a meaningful way because the word embeddings produced by Word2Vec tend to capture the _____ of words.

- semaphores
- syntaxes
- semantics
- syntactics

## Question 40 — 0.5 pts

"RNN" stands for...

- Rectified Neural Network
- Retrospective Neural Network
- Recurrent Neural Network

## Question 41 — 0.5 pts

"LSTM" stands for...

- ○ Long Short-tempered Memory
- ○ Long-shot Temporal Memory
- ○ Long Short-terminal Memory
- ○ Long Short-term Memory

## Question 42 — 0.5 pts

What's the amazing invention from 2017 that has catapulted Natural Language Processing (NLP) to great heights and has made present-day Large Language Models (LLMs) like ChatGPT possible?

- ○ Transformer
- ○ Transducer
- ○ Continuum Transfunctioner
- ○ Continuous Translator

## Question 43 — 0.5 pts

Attention is a key concept in Transformers (not the cartoon).

- ○ True
- ○ False