Write the set (1,2,3) as a row vector and a column vector.

rv = [1,2,3] cv = $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$

These vectors are ____transposes_____ of each other.

What is an instance in ML?

Data that we have about a real-world object/ an observation

What is a prediction in ML?

An estimated value for an instance, such as ham/spam

What is a target variable in ML?

a variable we want to make a prediction for

We use ____training data_____ to optimize an ML classifier. This is also referred to as _____labeled data_____ because the predicted values of the target variables are known.

The known values of these target variables are also referred to as _____labels_____.

Each type of data we have about an instance is called a _____feature/attribute_____ and we use these to base determine the predicted value for an instance.

We are observing dice rolls. Roll R can be any number from the set (1,2,3,4,5,6). The variable R is considered a _____random variable_____.

What are distance and similarity in ML?

Distance is a measure of difference between two instances, similarity is a measure of how alike two instances are

When a website recommends you a product based off of your purchase history, this is referred to as _____targeted advertising_____.

The process of picking the recommended product or products is referred to as _____collaborative filtering_____. What are the two versions of this process?

User-based collaborative filtering and item-based collaborative filtering

If a matrix contains mostly zeros, it is referred to as a _____sparse matrix_____.

In the problem 2^n = 16, n could be referred to as the _____exponent_____ or _____logarithm_____.

In that same problem, 2 would be referred to as the _____base_____.

What is the difference between supervised and unsupervised machine learning?

In supervised machine learning, labeled data is used to train the classifier. In unsupervised machine learning, the labels for the target variables are not known

Define the ML terms class, classifier, and classification.

A class is a possible value for a categorical random variable, a classifier is a ML program that performs the task of classification, classification is the task of assigning classes to instances based on the features of the instance

_____joint probability_____ is the probability that two events, A and B, both occur and is denoted by ___$P(A \cap B)$___.

The probability of these two events occurring at the same time might be _____conditional_____ if their joint occurrence is related, otherwise the two events might be _____independent events_____.

A collection of large texts (or documents) is known as a _____corpus_____. A unit of text within one of these documents is known as a _____token_____.

When analyzing text, we often omit words such as 'the' or 'and' because they do not contain useful information. We refer to these words as _____stop words_____.

When analyzing text, we often want to condense different tenses of a word (ex. 'run', 'runs', and 'running') into a single feature. To do this, we use a process called _____stemming_____.

One problem you might encounter when training a ML classifier is when a floating-point calculation returns an extremely small number. This problem is referred to as _____underflow_____.

A variable in a ML model that is set before the model is trained is known as a _____hyperparameter_____.

What is the difference between a positive and negative instance in ML?

A positive instance is of interest to us, while a negative instance is not

Define the ML terms true positive, true negative, false positive, and false negative.

True positive = we predicted positive and it was actually positive

True negative = we predicted negative and it was actually negative

False positive = we predicted positive but it was actually negative

False negative = we predicted negative but it was actually positive

Write the formula for accuracy using the variables TP, TN, FP, FN to represent the terms in the previous question. What is accuracy in simple terms?

(TP + TN)/(TP + TN + FP + FN)

Accuracy is, out of all our predictions, how many we were correct about

In order to determine which model best represents a set of data, we must determine the model's _____goodness of fit_____.

What is the formula for sample standard deviation?

$$\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})^2}$$

What does ddof stand for?

delta degrees of freedom

What is the difference between sample and population standard deviation?

With sample standard deviation, ddof = 1 (represented by 1/N-1 in the formula) meaning errors are being accounted for since the value is only representing a sample of the population. With population standard deviation, ddof = 0 (represented by just 1/N in the formula) meaning errors are not being accounted for since the value will represent the whole population

The variable used to represent a model is _____ ŷ_____, the variable used to represent a vector of labels is ___y_____, and the variable used to represent a matrix of instances and features is _____X_____.

Use the variables mentioned above to write the formula for finding residuals.

$$y - \hat{y}$$

What do SSE, MSE, and RMSE stand for? What are they used for when fitting a line/curve to some data?

SSE = sum of square errors

MSE = mean square error

RMSE = root mean square error

They are all measures of goodness of fit used to fine the line/curve that best fits some data

Y = ax + b is a function for a ____linear_____ regression, while y = ax^2 + bx + c is a function for a ____polynomial_____ regression.

What does OLS stand for?

Ordinary least squares

What is a statistic?

It is a number that describes some data

When analyzing climate data, the _____null hypothesis_____ would be that all variability in the trends in the data are simply caused by natural variability, and that there is no climate change.

_____p-value_____ tells us whether the probability of a statistic would be as or more extreme if the null hypothesis is true (or, in other words, that the calculation is statistically significant).

_____Pearson's r_____ tells us whether there is a correlation between two variables in a set of data.

_____r squared_____ is a measure of goodness of fit that, in a linear regression, represents the percent of variability among target variables that is explained by the predictor variables.

A function or functions that describes some data is referred to as a _____model_____.

_____optimization_____ is the task of finding the best solution from a set of possible solutions.

In ML, the task of finding the smallest distance between a model and a sample is known as _____minimization_____.

Observe the following functions and label them with the name of the value that they calculate:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

conditional probability

$$|A| = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = a \begin{vmatrix} \square & \square & \square \\ \square & e & f \\ \square & h & i \end{vmatrix} - b \begin{vmatrix} \square & \square & \square \\ d & \square & f \\ g & \square & i \end{vmatrix} + c \begin{vmatrix} \square & \square & \square \\ d & e & \square \\ g & h & \square \end{vmatrix}$$

$$= a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix}$$

$$= aei + bfg + cdh - ceg - bdi - afh.$$

Determinate of a matrix

$$\sum_{i=1}^{n} a_i b_i = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n$$

dot product

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{\det \mathbf{A}} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

inverse of a matrix

$$\sqrt{\frac{\sum_{t=1}^{T}(\hat{y}_t - y_t)^2}{T}}.$$

rmse

$$\sqrt{\sum_i (v_i - 0_i)^2} = \sqrt{\sum_i v_i^2}$$

magnitude of a vector

$$\boldsymbol{u} \cdot \boldsymbol{v} = \| \boldsymbol{u} \| \| \boldsymbol{v} \| cos\theta$$

dot product, geometric interpretation

For that last function, what do ||u|| and ||v|| represent?

Magnitude of vector u and magnitude of vector v

For the following sets of conditions, identify the special matrix they are describing:

N = m _____square matrix_____

a[i][j] == 0 and i != j _____diagonal matrix_____

a[i][j] == 0 and i > j _____upper triangular matrix_____

a[i][j] == 0 and i < j _____lower triangular matrix_____

(a[i][j] == 0 and i != j) and (a[i][j] == 1 and i == j) _____identity matrix_____

a[i][j] == a[j][i] _____symmetric matrix_____