Jeopardy 7 Question Sources:

**Correlated Data:**

1. What is the assumption that is violated when dealing with correlated data?
   1. The most common assumption we’ll make in modeling is that our observations are independent of one another.

2. What are 3 types of correlated data and give an example of each:

timeseries - stock prices, number of items sold over time

  spatial - voting patterns - some sort of spatial relationship

  network data -

3. What is omitted variable bias:

if your models leaves our important predictors, the bias that occurs is known as omitted variable bias (the bias that exists because we omitted, or left out, and import predictor)

1. What is the expanding mean?

The "expanding mean" simply uses all data points up to the current time to calculate the mean, as opposed to a moving window. the average of that row and every row before it.

1. is the difference between expanding mean and exponentially weighted windows?

Exponentially weighted windows are one of the most common and effective ways of averaging out noise in time series data. The averaging is done with an "exponential decay" on the contribution of prior means, decreasing the contribution of time points that are further in the past.

The (adjusted) exponentially weighted mean for time, tt, is defined as:

TimeSeries:

1. What do you lose when you call .diff or .pct\_change() and why is this good for modeling?

you lose the trend. #you lose the trend when you change it to the percent change or find the difference

# this is good for the model - the mdoels can't see trends, so it's important to take that out of the equation - you need to get that out using .diff or percent change/

2. . Describe what and ACF and PACF show us:

ACF= Specifically, autocorrelation measures how closely earlier values relate to values that occur later in time.

#a way to quantify whether or not observations are correlated with each other

# autocorrelation is the correlation between a series and itself at different lags

PACF= This is similar to autocorrelation, but, instead of being just the correlation at increasing lags, it is the correlation at a given lag, *controlling for the effect of previous lags/timepoints.*

3.suppose we want to predict the price of bitcoin over time. The price has long term trends, but is also susceptible to sharp/sudden increases or decreases in price.

4. what do parameters p, q and d represent?

#p the number of previous time points that we want to incorporate into our model

#q As with autoregressive models, we have an order term and we refer to our model MA(q) – the moving average model relies on the last q errors – so q is the last number of errors you want to include

d = difference – remember when we manually found the best difference manually in the bus\_ridership lesson. We performed the dickey fuller test to find the difference that gives us stationarity, then used that in the arima model.

5. What is stationarity and how can we check for it?

**Stationarity**

A **stationary** time series is one that:

* has a constant mean over time.
* has an be.)

To check stationarity:

* Visually look at the mean. Is it constant? If not, then our time series is not stationary.
* Visually look at the autocorrelation. Does the relationship between, for example,  the same as the relationship between  not, theour time series is not stationary.
* Dickey-Fuller test.

6. How do you interpret the p-value of 0.134 when performing a dickey fuller test?

how to interpret the p value:

# because our p value is greater than alpha (our signficance level, 0.05, we fail to reject ou null hypothesis null hypothesis - is that our data are not stationary). we cannot conclude that our data are stationary.

**GeoSpatial Data:**

1. What is Tobler’s First Law of Geography?
   * Everything is related to everything else, but near things are more related than distant things”
2. In GIS, what is W?
   * our weight matrix - it accounts for the dependencies in our observations

We use a weight matrix ! to measure how closely related two regions are. This describes how much weight to assign to region " when measuring region #.

1. Given spatial data, How might we create a weight matrix?

1 if region # borders region " (or is within a certain distance of "), otherwise 0.

Use the inverse distance between centroids or between landmarks

4. When working with spatial data – what would be an example of an areal process?

If our spatial domain D is a set of non-overlapping regions, then we are working with an areal process. (i.e. states, ZIP codes)

• what would be an example of a point pattern process?

5. If our spatial domain D is continuous (likely 2-D or 3-D space), then we are working with a geostatistical process. (i.e. rainfall)

• If our spatial domain D is a collection of random points, then we are working with a point pattern process. (i.e. locations of car accidents)

**Network Analysis:**

1.What are 4 or the 5 types of networks we discussed in lecture and give an example of each:

types of networks:

   undirected - is one where the connection extends in both directions - the relationship between a and b is the same as the relationship between b and a

                - examples:

                        facebook friends

                        sexual partners

                        family trees

                        (if i am connected to you, you are connected to me)

    directions - a directed network is one where the connection may only flow in one direction

                -examples:

                    twitter/instagram - you can follow someone without them following you

                    payment/venmo - payment may only flow in one direction

                    food chain

                    disease transmission

    cyclic - contains at least one cycle. a cycle exists when a node can be connecte to itself by traversing at least one edge

            - circle of friends, circle of life/food chain

    acyclic: a network that has no cycles

            - corporate organization charts (node - employee, edge - supervisor relationship)

            decision trees

            family tree - depends on how edges are defined

    multigraph - may have multiple links connecting the same pair of nodes

        - royal family tree

        - choices of flights between cities

        -travel routes

1. tell me how many degrees exist for each node in this network - specific in degrees and out degrees:

degree- the degree is a node in an indirected graph is given by the number of links of that node in-degree - the number of links heading into a node out-degree - the number of links heading out of that node

3. often, what does the distribution of a network look like?

* right skewed

1. build an adjacency matrix for the following network:

5. what does it mean if a two graphs are isomorphic. draw an example of this:

we might say two graphs are isomorphic to one another if we can twist one graph to look exactly like the other without cutting or gluing.

        - all of their properties are the same

6. what are the three main difficulties with network analysis:

difficulties with network analysis:

 storage

 inference - we take a sample from the network - the degrees/connections may be underrepresented because we're not looking at the whole network' - when we sample nodes or edges, we run into problems -it's difficult to calculate the average degree based on a sample’

 vizualization - nodes can be drawn in a variety of ways, which makes vizualizing networks complicated.

**Review:**

1. What is a loss function and give 3 examples some examples?

* + Some measurement of error that we want to minimize
  + Measure of model performance

Something that measures how bad our model is

* MSE – mean squared error
* L1 & L2 (lasso and ridge regularization)
* Sse +lambda\*beta
* Gini
* Entropy
* Categorical cross entropy – might want to optimize a categorical cross entropy in a classification problem

2. What purpose does a loss function serve?

We want to find the values of our parameters that minimize our loss function (basically, minimizes the crappiness of our model – minimizes the error)

3. Suppose you’re using a cart model to predict if someone is going to go to a Halloween party or not. We classify y as yes – will go to a Halloween party, no, will not go to a Halloween party. Our taining set has 60 Y and 60 N. the algothim needs to decide which variable to split on first, F1 or F2. Which would be the better choice, and why?

What is the purpose of the kernel in support vector machines?

* + To create linearly separate boundaries in our data

Gradient Descent:

4. What is alpha in gradient descent?

The learning rate – controls how fast we move with each step. We’re taking these steps down the gradient to optimize our loss function – alpha says how far we’re going to go for each step. The larger alpha is, the faster we’re going to learn.

5. What is a pitfall of having a high learning rate, what about a low learning rate?

The algorithm might skip over the optimal values for the parameters. – may never converge!

Low learning rate – it could take forever to run.

Final Jeopardy:

In general terms (don’t need to get super mathy – just focus on the concept):

What is gradient descent and how does it work and what is it’s goal?

Gradient descent is

An iterative method

Used to identify the optimal value of parameterers

By optimizing an objective function – minimize loss function

Start by making a guess for the optimal parameter value

Calculate the loss given the parameter value

Update guess to decrease loss

Keep going ntil loss is sufficiently minimized

Goal:

Find the best possible values for our betas

Derivative:

The slope of the curve

Gradient loss function with respect to b1

Tells us direction of the steepest slope

For each step, we go in the direction of the steepest step.

Reorient at each step

Greedy algorithm –just trying to optimize

Keep going until-

We hope we’re at the minimum here

Slope is close to zero

When the slope is zero, it means that the algorithm has converged.

1. instantiate model
2. select a learning rate
3. select a starting point
4. calculate the gradient loss function to minimize the loss function as quickly as possible
   1. best direction to head into as it relates to beta 1
5. calculate beta1 i+1
6. c

want to find the slope (derivative) we want to decrease our loss function