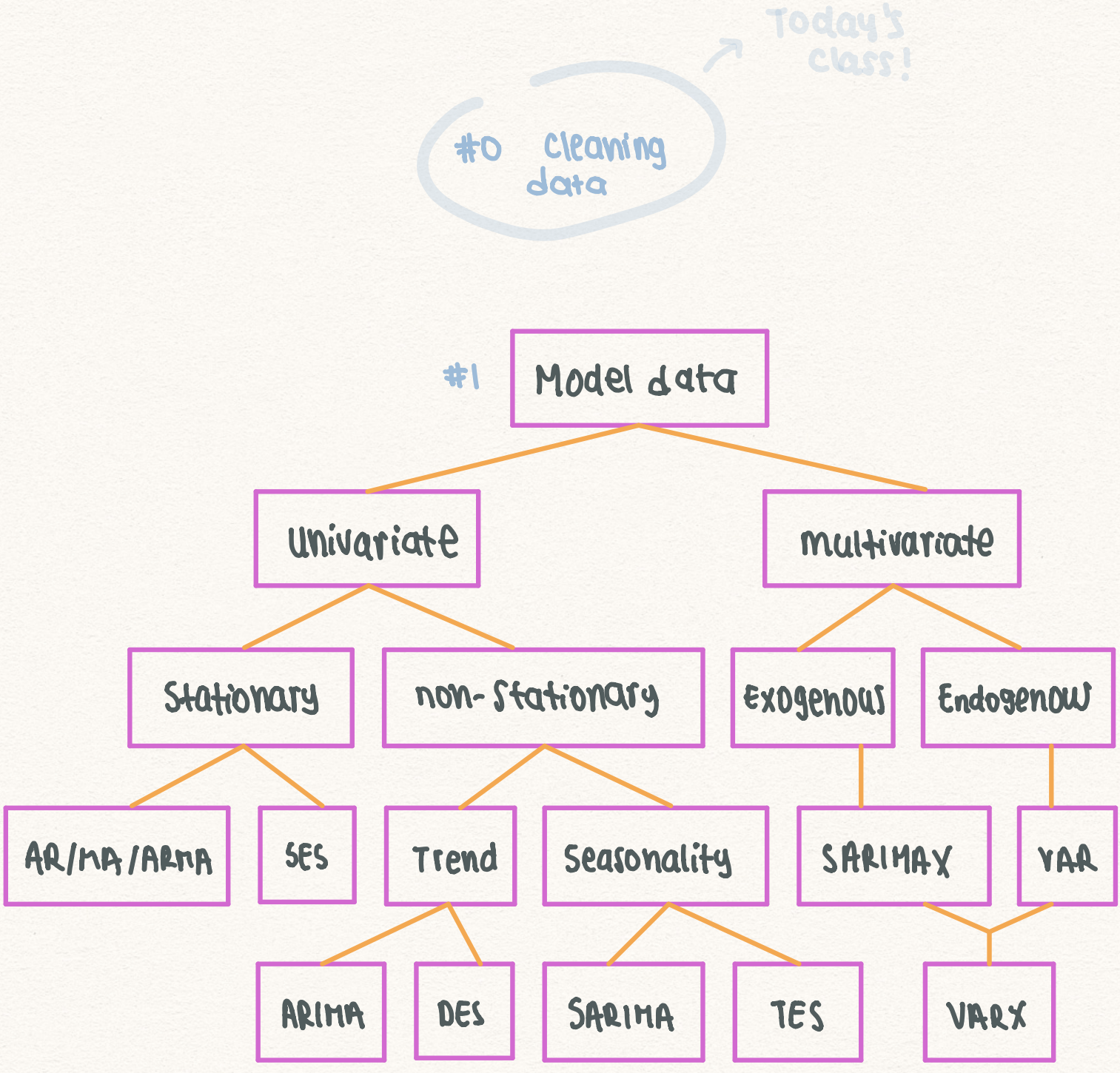
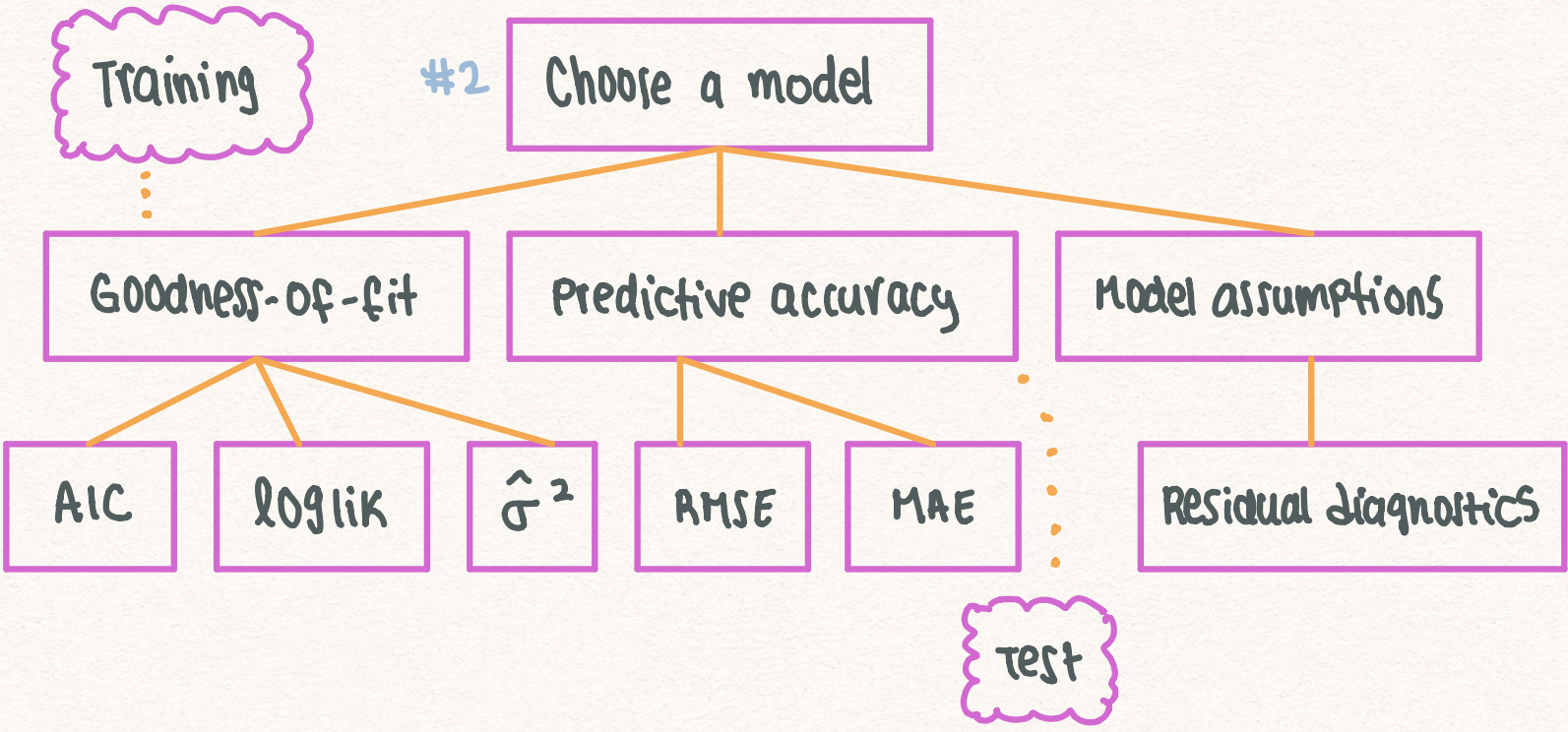
Date: Tuesday, December 4, 2018

Big picture




Training
#2
Choose a model
Goodness-of-fit
Predictive accuracy
Model assumptions
AIC
loglik
$\hat{\sigma}^2$
RMSE
MAE
Residual diagnostics
test

#3 Forecast

Imputing Time Series

As with any modeling endeavor, time series often has missing observations. In order to fit our models we require complete data. For this, we apply methods of imputation.

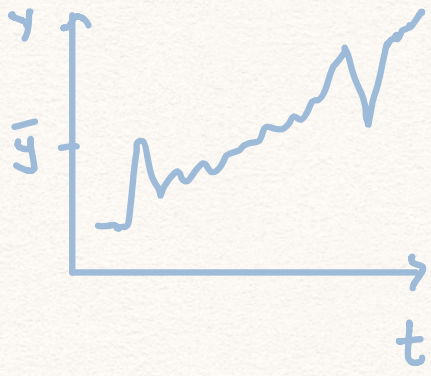
* Many general methods of imputation exist, but often they do not account for the time and correlation structure that you need to consider when dealing with time series data.

* Effective methods of imputation account for the time and correlation structure of the data. The specific methods used will depend on whether the data is stationary or if it has trend and/or seasonality.

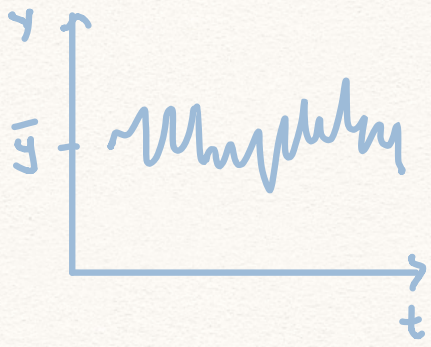
NON-TS-Specific methods

- Mean imputation
- Mode imputation
- Media imputation

} Fill gaps with a measure of center calculated from the observed data.



* these don't work well in the presence of trend.



* these can work well if your data is stationary.

- Random-sample imputation } Draw a random observation from the existing data and use that to fill gaps.
  → You can tune it to intervals.

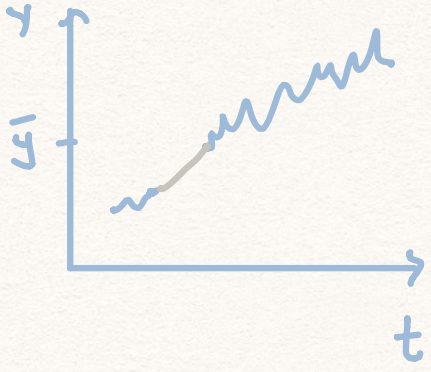
  ↳ For the same reasons as above, this may be appropriate for stationary data, but not otherwise.

TS-specific methods

- Last observation carried forward (LOCF)
- Next observation carried backward (NOCB) } These exploit the existence of strong correlation.

  * Less effective for filling big gaps.
  * It will not work well if adjacent observations are very different. This may happen when we have strong effects.

- Interpolation (linear or polynomial)



  → Fit a straight line accross the gaps in the data or higher polynomials if the data warrant it.

- Seasonal adjustment + interpolation

  → De-seasonalize the data (which can be done in the presence of missing data).

  → Perform interpolation on missing observations.

  → Once missing data has been imputed, re-seasonalize it.

[Break]

[R]

Clustering time series

In the context of time series, we may use Clustering to identify common patterns and shapes and group entire series accordingly.

Example: It may be of interest to group different retail products on the basis of similar sales patterns. Or you may want to apply clustering to a large corpus of times to determine which could effectively be used in a multiplicative time series model.

Partitional clustering is performed using the k-medoids algorithm which behaves exactly like k-means except that the centroids are actual time series "prototypes" rather than the average of all of the time series.

(1) Dynamic time warping!

[A]