

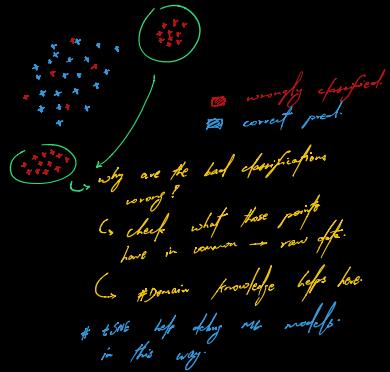
- # Single feature weights can lead to major decisions in business
No fear to overexplanations &
- # Single is better
No need to write a fucking thesis full of charts
- In the end, only the weights matter. Results, not the fucking overexplained analysis.

Visualizing High Dim Data

PCA → linear
t-SNE → non-linear neighborhood.

HIGH DIM \rightarrow t-SNE \rightarrow LOW DIM

No need to make a model.
first see the effects



- # Model based
 - if separable
 - single ones
 - * up hand, (DT) is very interpretable
 - * get feature importance for words \rightarrow how more words \rightarrow see which words were given more importance
does it look logical?

get features somehow &
choose some of something from
chapter. see if something finds of
gives an idea if
a feature selection technique
is working

PART - 2

* Asking the right questions

* Always look at the raw data
& not just aggregate numbers

* High level stats

- # No. of different vs features
- # Label imbalance
- # Distribution of y_i for regression

* Feature wise analysis

- # categorical \rightarrow distribution of few
- # real value \rightarrow distribution of lots
- # missing values & imputation
- # outliers & their impact

* What other features might work well?

- # group by + count on raw off \rightarrow # of times someone bought something
- # Binary features using DT \rightarrow ~~int age cat1 cat2~~ \rightarrow find DT
- # See how DT has done the work & know the information of features.
- # Transform features \rightarrow log, exp, sqrt, \sim , bin-cos \rightarrow always better as non linear
- # Matrix transformation based features \rightarrow in the worst I forgot most transforms work better if it's general.
- # Autoencoder ~~DT~~
- # Clustering of features \rightarrow add new feature "cluster-no"
- # Different encoding \rightarrow TF-IDF w.r.t etc

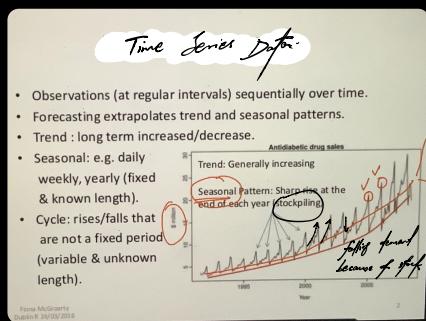
- # Feature vs output
- * correlation \rightarrow with only good linear/multifac. relation
- # $p(x_i = 1 | F_{ij} = h)$
- # Build a model with just one feature \rightarrow can find complex relations



overlap is inevitable, this is ok.
if there's no overlap, there's no need for multiple models.

- # High Dim \hookrightarrow
 - # Are the bag points clustered?
 - # Difficulty specifying mapping spaces?
 - # Visualize that is that performs best.
 - Affine invariant examples -> add off
- } add features

Time Series:
logistics, Fourier transform, moving averages. \rightarrow NYC Taxi Case Study.



All of this is in the world,
this is just a typical scenario.

Shared Entry Recognition
↳ figure out a name in
a sentence
Bi-directional LSTM, Transformer
as state of the art. And no one
finishing knows how exactly they
factor so well.

Image Data \rightarrow CNN Segmentation
+ + SNE
+ look at raw images

imagine f_{ij} & g_{ij} are categorical model

\hookrightarrow joint later with (χ^2)

\hookrightarrow just build a (model)

can model complex relationships

only defines linear or monotonic relationships

if f_{ij} is enormous, why
bother with χ^2 . χ^2 a old school
stuff was built for small data.

is very slow
shortcut AF
with so much plotting, very is
unusable

Research & me cannot be separated.
idea driven can't do fulfilling grants in me. It's full
of fakes with rights who expected me to waterfall
time. with me projects.

The way of trying a organization
is too slow, so don't try to have
the best. That is anything.

It by Ian Groves
Judas begins up
Good one, but
affected.

There are just steps, how to
use them? "it depends"

↓
depends on what?
how to know that?

⇒ PRACTICE a key

it's like knowing keys on a
piano. A to G.

How to play Beethoven?
fulfill practice.