Additional Topics

Jayanth Rasamsetti
CEO & Founder www.sgmoid.com
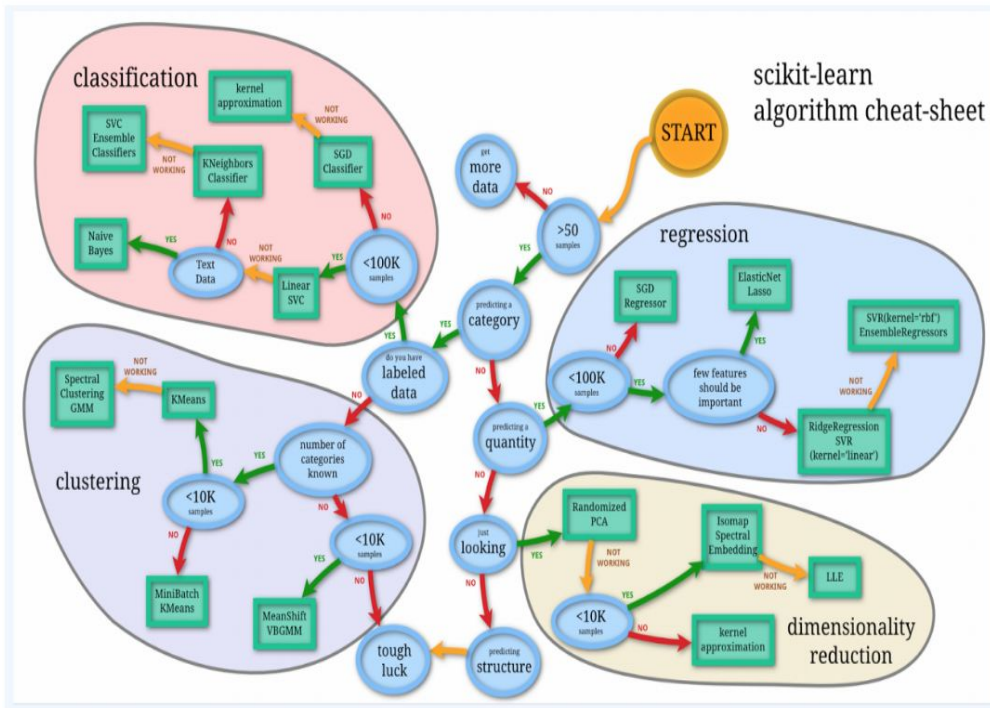Columbia University
IIT Madras (B.Tech & M.Tech)

Remember the flow

"Data Science is a process, not a result.

Data Science is minority rule.

Even if 300 Machine Learning winners agree on something, one experiment can prove them wrong."

Remember the checklist

Import matplotlib line at the beginning (Always!). Import all necessary libraries.

Read the train and test data, in case of large datasets, create a subset in a nearby path

Data Preprocessing:
a) Good idea to view the head
b) Check the shapes of train and test
c) Check dtypes()
d) Check for missing values (Impute them)
e) Convert categorical variables into numericals (using get_dummies)
e) Scale train, val and test (zscore)
f) Check for class imbalance (handle using sampling, smote, etc)

In theory, there's no difference between theory and practice. In practice, there is.
   *-   Alex Hardy*
10,000 hour rule

Remember the checklist

Exploratory Data Analysis (EDA)

Sketch plots between classes (class imbalance), or in case of regression the entire continuous variable

Sketch Correlation Plots between all possible features (can we drop variables or feature engineer?)

Sketch box plots for the all variables Sketch plots of class distribution for simple features e.g. class vs age class vs education class vs income etc
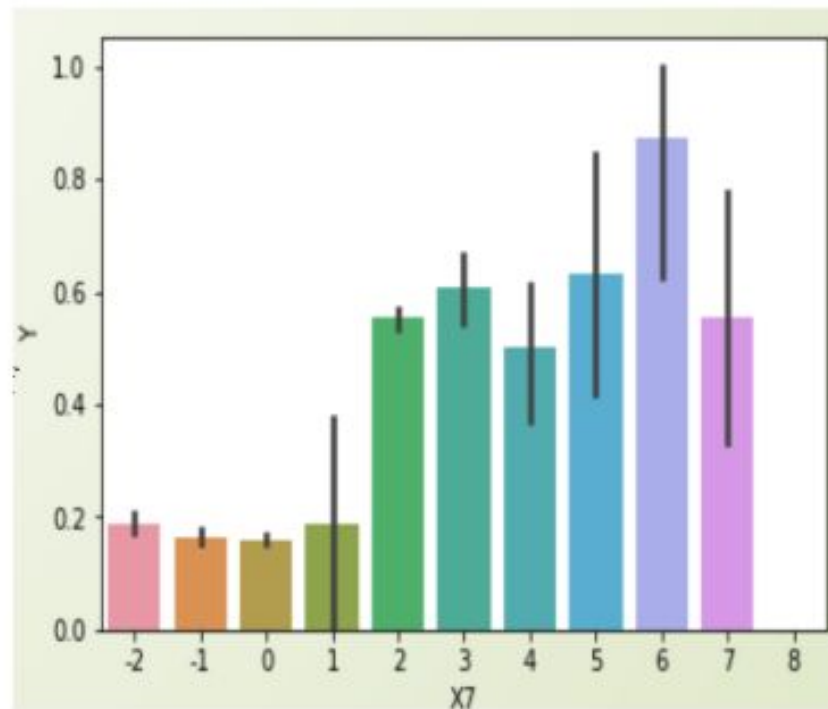
Recap & Next Steps

We have learned and understood:

Supervised Learning (Decision Trees, KNN, Naive Bayes)
Parameter Tuning

Unsupervised Learning (HAC, K Means)

[Practiced on 8 different datasets & several more Case Studies]

Best Practices

1. Listen attentively
2. Code during most productive hours (i.e. in the morning, after a run, etc)
   Do not code right after lunch, just before bed, etc
3. Get strong peers *(you are the average of the 5 best people next to you)*
4. Stick with positive people *(be a roach).* Positive people will see that you are tenacious and will appreciate that you are seeking guidance
5. Practice 30 minutes of python code everyday. Write your own helper functions. Upload a submission to Kaggle everyday and write kernels

Most valuable skills in the world: Sales, Tech (You can pick up on the job finances & legal, or learn through an experts eyes)

Your mind

3 phases

1. **Limbo Phase:**
2. **Discovery Phase:**
3. **Growth Phase:**

Limbo Phase:

1. You are generally sleepwalking. Difficult to get out of bed & run through your tasks
2. Default reaction is to blame others [90% confusion, 5% indecision]. Highly impatient
3. Short bursts of hours for instant gratification. Flighty mindset and low confidence in abilities (but oozes out as arrogance or strong opinions)
4. During an event unable to discern between your mistake and others mistakes. Lot of overthink.
5. You know you are destined for greater but unable to generate the escape velocity

Your mind

Discovery Phase:

1. You start earning accolades, awards and some recognition. Your esteem is back
2. Your mind is spending longer hours than growth phase, you start losing track of when you are hitting the sack. You keep an alarm to wake up at 7 am.
3. More self doubt about skills. You start avoiding negative people.
4. You complete a lot of the todo lists. You are still struggling to complete your todo lists, [70% confusion, 30% clarity]. Going with the flow.
5. You are still succumbing to your key personality traits.

Your mind

Growth Phase:

1. Listening far more attentively. You are able to manage all tough relationships (parents, siblings, spouse, etc). You have a high degree of forward momentum.
2. You value time (others time and hence your own time). Stronger 'first principles' outlook and ability to think in a 'meta' way.
3. Less self doubt about skills, more self doubt about purpose
4. Able to fairly keep the todo list in control, but still forget some items
5. Caveat: You will still have a lot of residual properties from each phase

Goal

Your goal should be to first identify which phase you are in and then transition to the next. To generate escape velocity, you need energy and positive momentum.

Chose a career which has 3 things: 1) You love 2) Strong vested interest 3) Something you will stick to for 7 years

Naval @naval · Nov 17

In physical domains (diet, workouts), consistency is king.

In intellectual and social domains, wins are rare, sudden, and nonlinear - 99% of effort is "wasted."

Once we find our ideal spouse, career, friends, we realize how much time we squandered by not moving on sooner.

💬 104   ↻ 2.2K   ♡ 8.8K   ✉