

INTRO TO DATA SCIENCE

SESSION 20: WHERE TO GO NEXT?

I. WHERE HAVE WE BEEN?

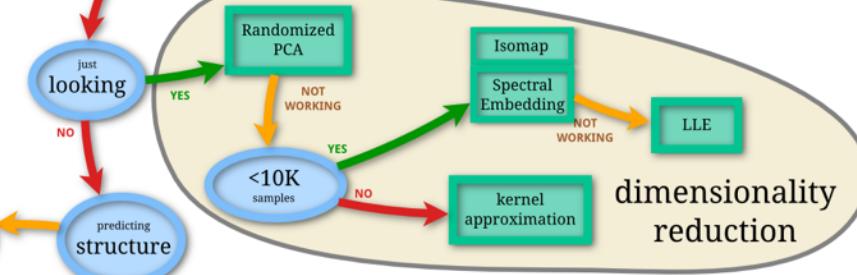
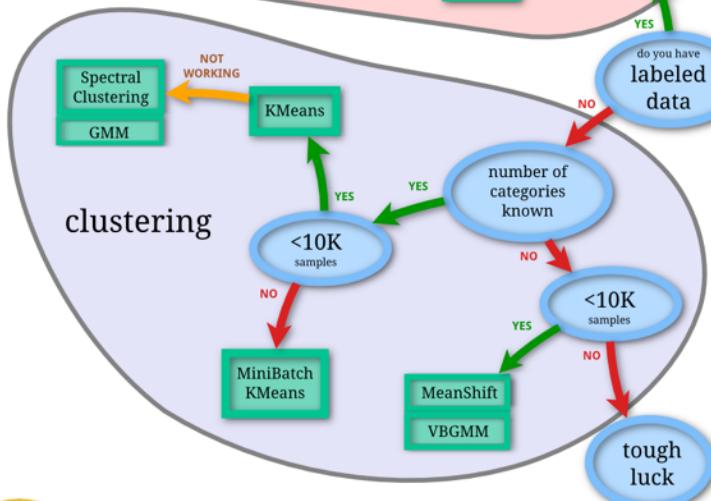
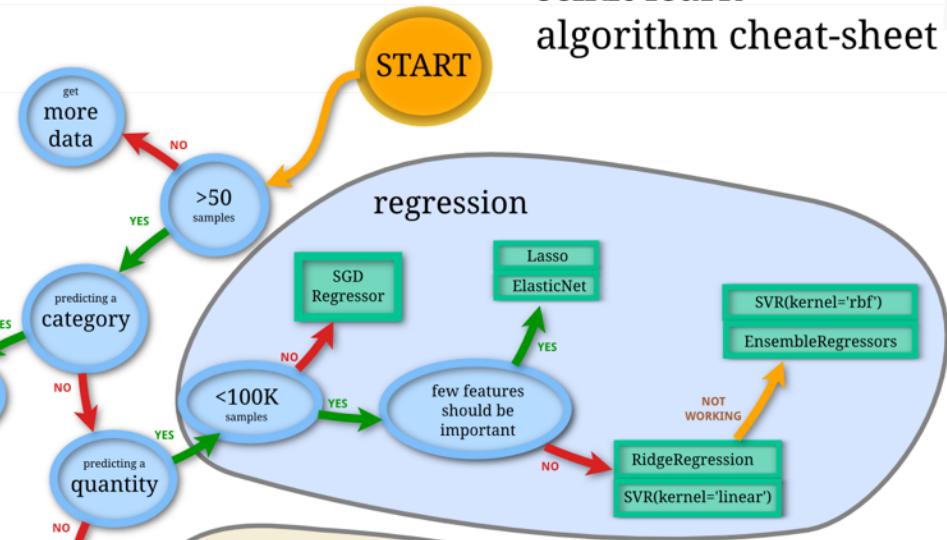
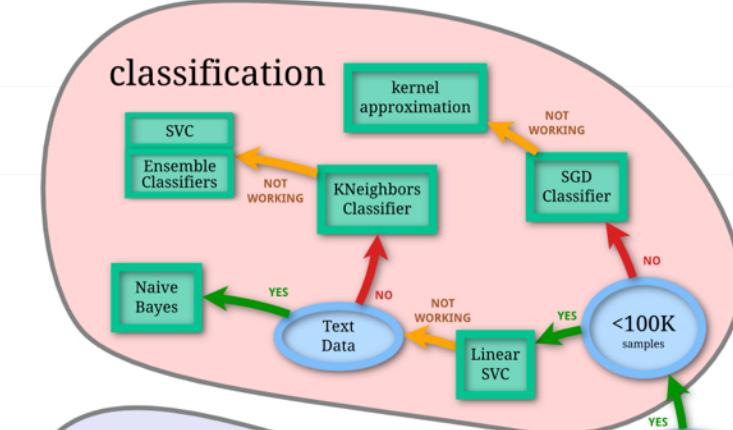
	<i>continuous</i>	<i>categorical</i>
<i>supervised</i>	<i>regression</i>	<i>classification</i>
<i>unsupervised</i>	<i>dimension reduction</i>	<i>clustering</i>

WE COVERED

- *Data acquisition and preparation*
- *Exploratory data analysis*
- *Supervised learning:*
 - *kNN*
 - *Linear and multiple regression*
 - *Decision trees & random forests*
 - *logistic regression*
 - *Naïve Bayes*

- *Unsupervised learning:*
 - *K-means clustering*
 - *PCA/SVD for dimensionality reduction*
- *Text mining and natural language processing*
- *Support Vector Machines (SVM)*
- *Recommender systems*
- *Overview of time series*

scikit-learn algorithm cheat-sheet

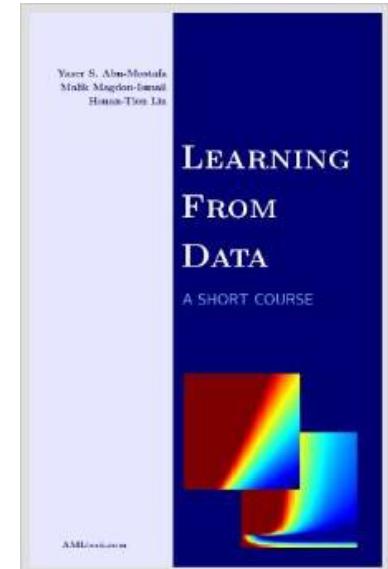
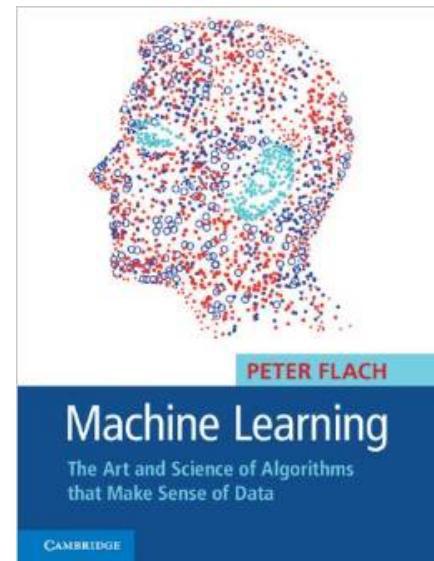
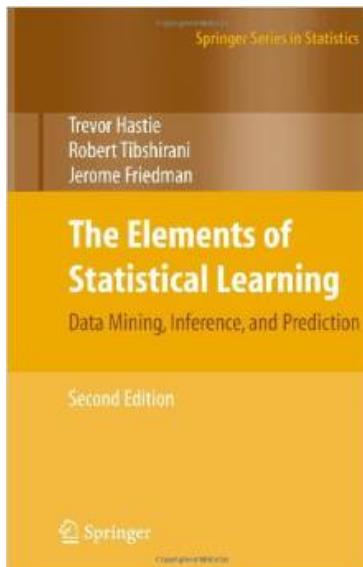
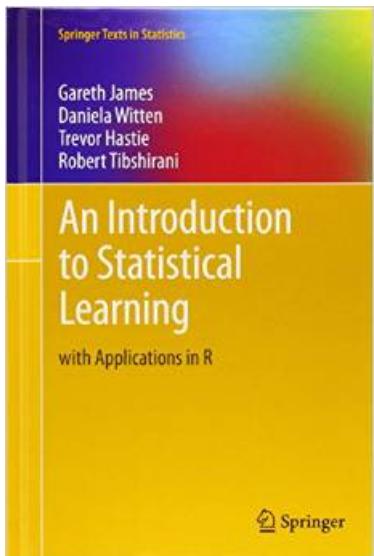


WE USED

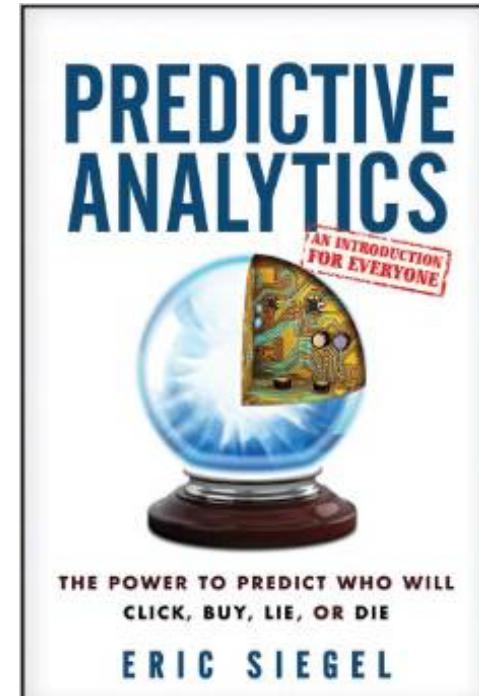
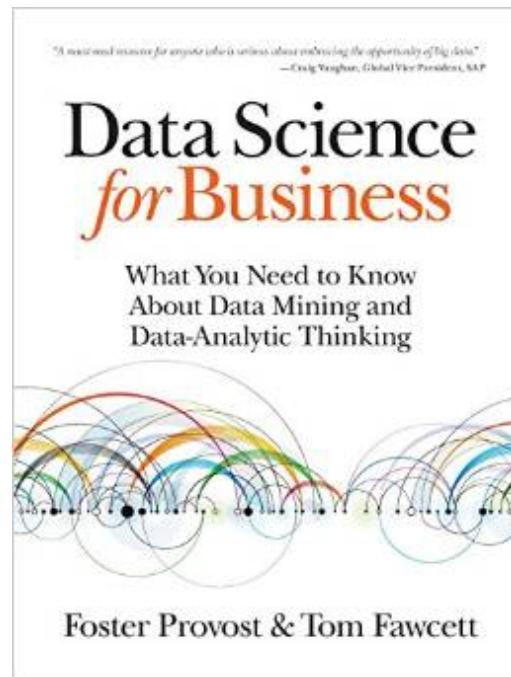
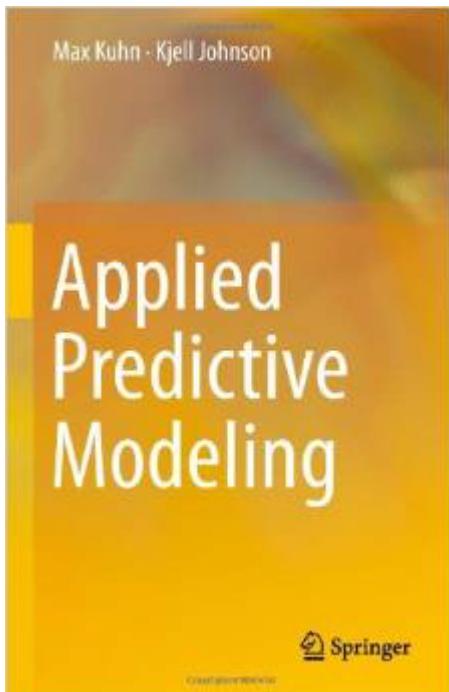
- *Python and iPython Notebook*
- *Scikit-learn and Pandas*
- *matplotlib and Seaborn*
- *NLP*
- *SQL and SQLite*
- *Python recsys*

II. WHERE TO GO NEXT?

WHAT TO DO NEXT: READING



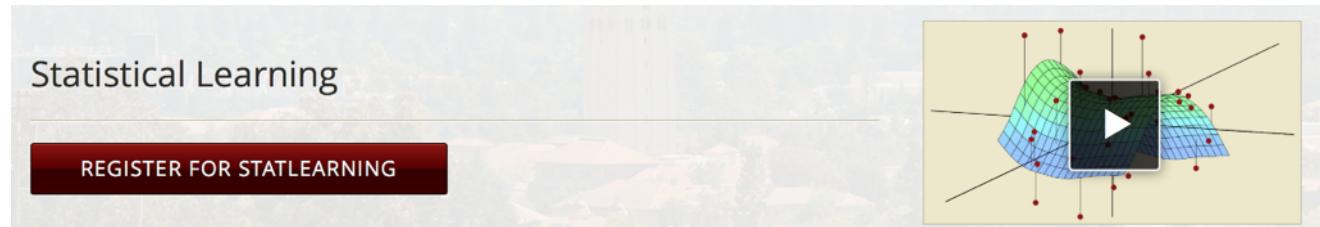
WHAT TO DO NEXT: READING



WHAT TO DO NEXT: READING



WHAT TO DO NEXT: ONLINE COURSES



Stanford University
Machine Learning

Ended 3 years ago

[Course Record](#)

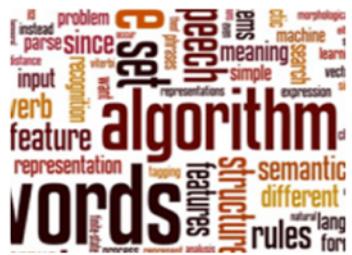


Stanford University
Mining Massive Datasets

Ended 2 months ago

[Course Record](#)

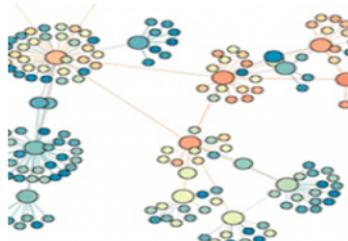
WHAT TO DO NEXT: ONLINE COURSES



Stanford University
Natural Language Processing

Ended 3 years ago

[Course Record](#)

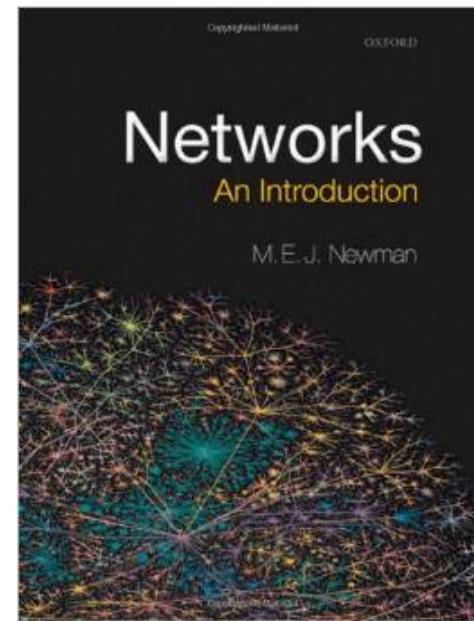
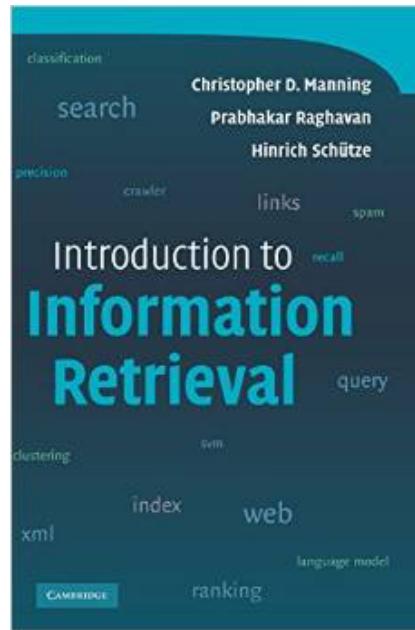
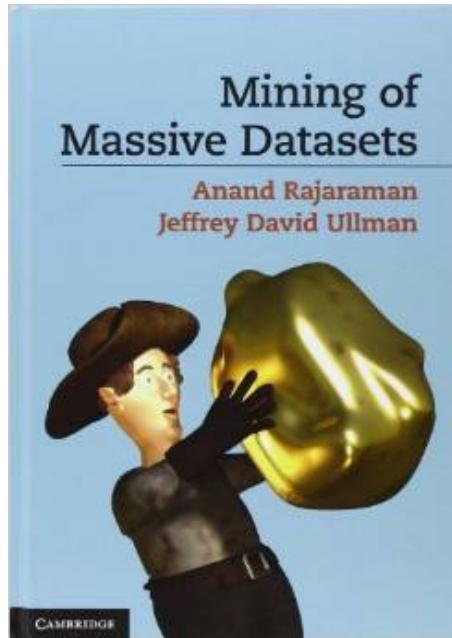


University of Michigan
Social Network Analysis

Ended a year ago

[Course Record](#)

WHAT TO DO NEXT: READING



WHAT TO DO NEXT: STACKOVERFLOW PRO TIP

The image shows two side-by-side screenshots of the Stack Overflow user profile for 'ogrissel'. The left screenshot shows the main profile page with sections for Profile, Activity, Communities, and Top Tags. The right screenshot shows the Network Profile page with sections for Top Network Posts, Top Posts, and Badges.

User Profile (Left Screenshot):

- Profile:** Shows a cartoon portrait of a man with a beard, 15,905 reputation, and activity counts for answers (2), questions (35), and badges (52).
- Activity:** Shows a timeline of recent activity.
- Communities (10):** Shows participation in Stack Overflow, Cross Validated, Area 51, Meta Stack Exchange, and Server Fault.
- Top Tags (182):** Shows the top tags: scikit-learn (1,008 posts), python (912 posts), machine-learning (676 posts), svm (556 posts), classification (142 posts), and scikits (133 posts).

Network Profile (Right Screenshot):

- Top Network Posts:** Shows the top posts from other Stack Exchange sites.
 - What are common statistical sins? (75 votes, Python)
 - Machine Learning using Python (58 votes)
 - Choice of K in K-Fold cross validation (54 votes)
 - Large scale text classification (34 votes)
 - Is there a concept of "enough" data for training statistical models? (25 votes)
 - LARS vs coordinate descent for the lasso (23 votes)
 - What is the objective Scikit-learn's Random Forest classifier is optimizing at each node? (22 votes)
 - Is scikit-learn suitable for big data tasks? (21 votes)
 - What's the difference between LibSVM and LibLinear (18 votes)
 - How to elementwise-multiply a scipy.sparse matrix by a broadcasted dense 1d array? (17 votes)
- Top Posts (338):** Shows the top posts on Stack Overflow.
 - Is there a recommended package for machine learning in Python? (75 votes, sep 6 '10)
 - Save NaiveBayes classifier to disk in Scikits learn (58 votes, jun 23 '12)
 - Python: tf-idf-cosine: to find document similarity (54 votes, aug 26 '12)
 - Python scikit-learn: exporting trained classifier (34 votes, jul 7 '13)
 - Classifying Documents into Categories (25 votes, jun 24 '10)
 - Possibility to apply online algorithms on big data files with sklearn? (23 votes, sep 17 '12)
 - Is scikit-learn suitable for big data tasks? (22 votes, jun 10 '13)
 - What's the difference between LibSVM and LibLinear (21 votes, jul 17 '12)
 - How to elementwise-multiply a scipy.sparse matrix by a broadcasted dense 1d array? (18 votes, jul 14 '10)
 - fattest SVM implementation usable in python (17 votes, feb 15 '12)
- Badges (89):** Shows the user's badge collection.

<http://stackoverflow.com/users/163740/ogrissel>

WHAT TO DO NEXT: TECHNICAL SKILLS

- *Python: use it every day / week!*
- *Scala (if you get bored with Python; Spark has PySpark)*
- *Visualization (D3 and js)*
- *Algorithms: Artificial Neural Networks / Deep Learning*
- *“Big Data:”*
 - *MapReduce / Hadoop (?)*
 - *Spark*
 - *Elasticsearch*

kaggle



[topcoder]TM

DataKind

WHAT TO DO NEXT: MEETUPS

SF Machine Learning

<http://www.meetup.com/sfmachinelearning/>



<http://www.meetup.com/Data-Mining>



Data Visualization Group in the Bay Area

<http://www.meetup.com/visualizemydata>

WHAT TO DO NEXT: ONLINE

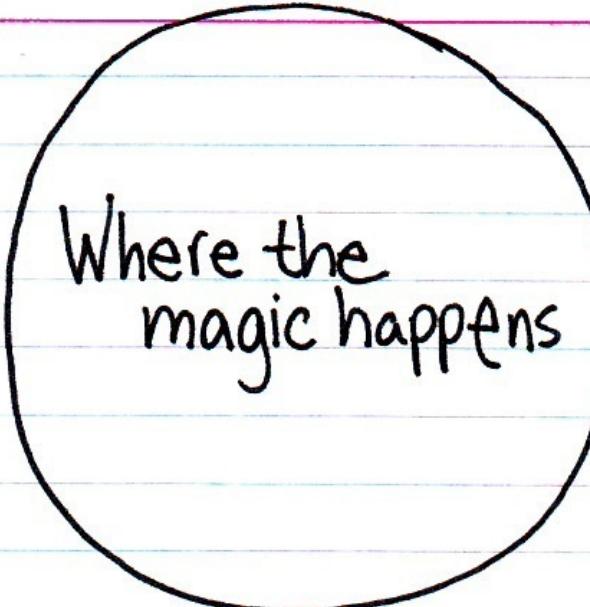
- *Online :*
 - <http://www.reddit.com/r/machinelearning>
 - <http://www.r-bloggers.com>
 - <http://dataelixir.com>
 - <http://www.datatau.com>
 - <http://pydata.org>
 - <http://www.technologyreview.com>
- *LinkedIn groups on ML and DS*

Each Other!!!

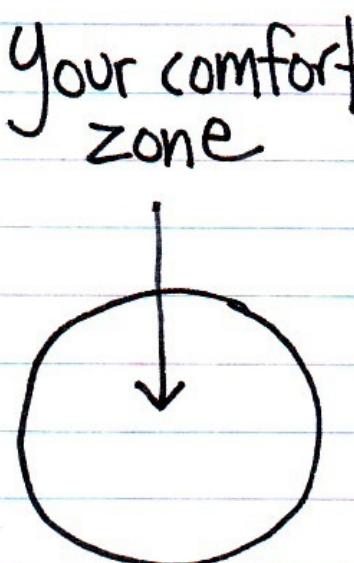
- *Class Google group / mailing list*
- *LinkedIn*
- *Study groups (DAT3 still does dinner and shares job opportunities)*

CONGRATULATIONS...

...for getting out of your comfort zone!



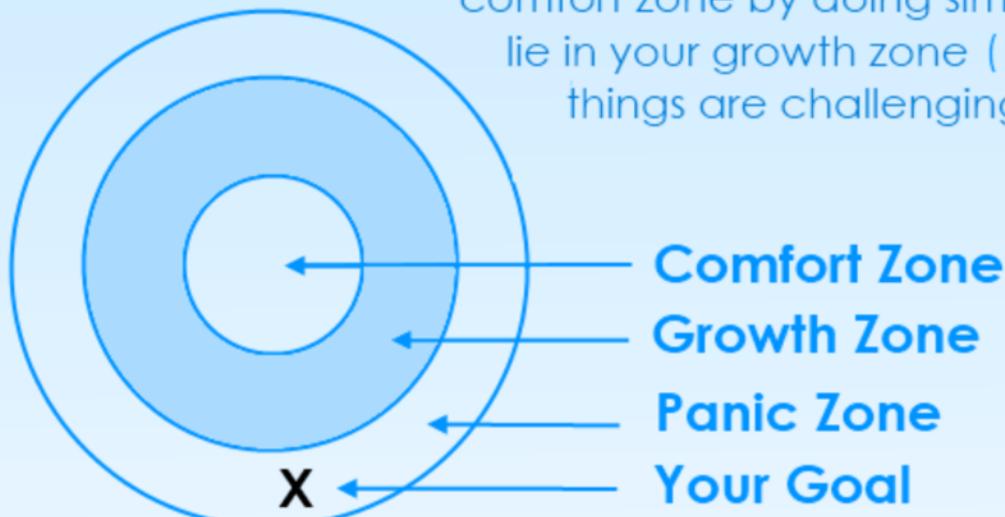
Where the
magic happens



Your comfort
zone

How to Grow Your Comfort Zone

Any goal or challenge may fall into one of three zones - your comfort zone, growth zone, or panic zone. If your goal is currently in your panic zone, i.e. it would be too scary to do now, you will need to grow your comfort zone by doing similar challenges that lie in your growth zone (the zone in which things are challenging or scary, but do-able).



How to Grow Your Comfort Zone

As you pursue challenges in your growth zone, those challenges become easier and your comfort zone expands.

Eventually, challenges that were previously in your panic zone begin to fall into your growth zone, and ultimately within your comfort zone.

