GENERAL ASSEMBLY

INTRO TO DATA SCIENCE

SESSION 20: WHERE TO GO NEXT?

DAT13 SF // May 13, 2015

I. WHERE HAVE WE BEEN?

REMEMBER THIS…?

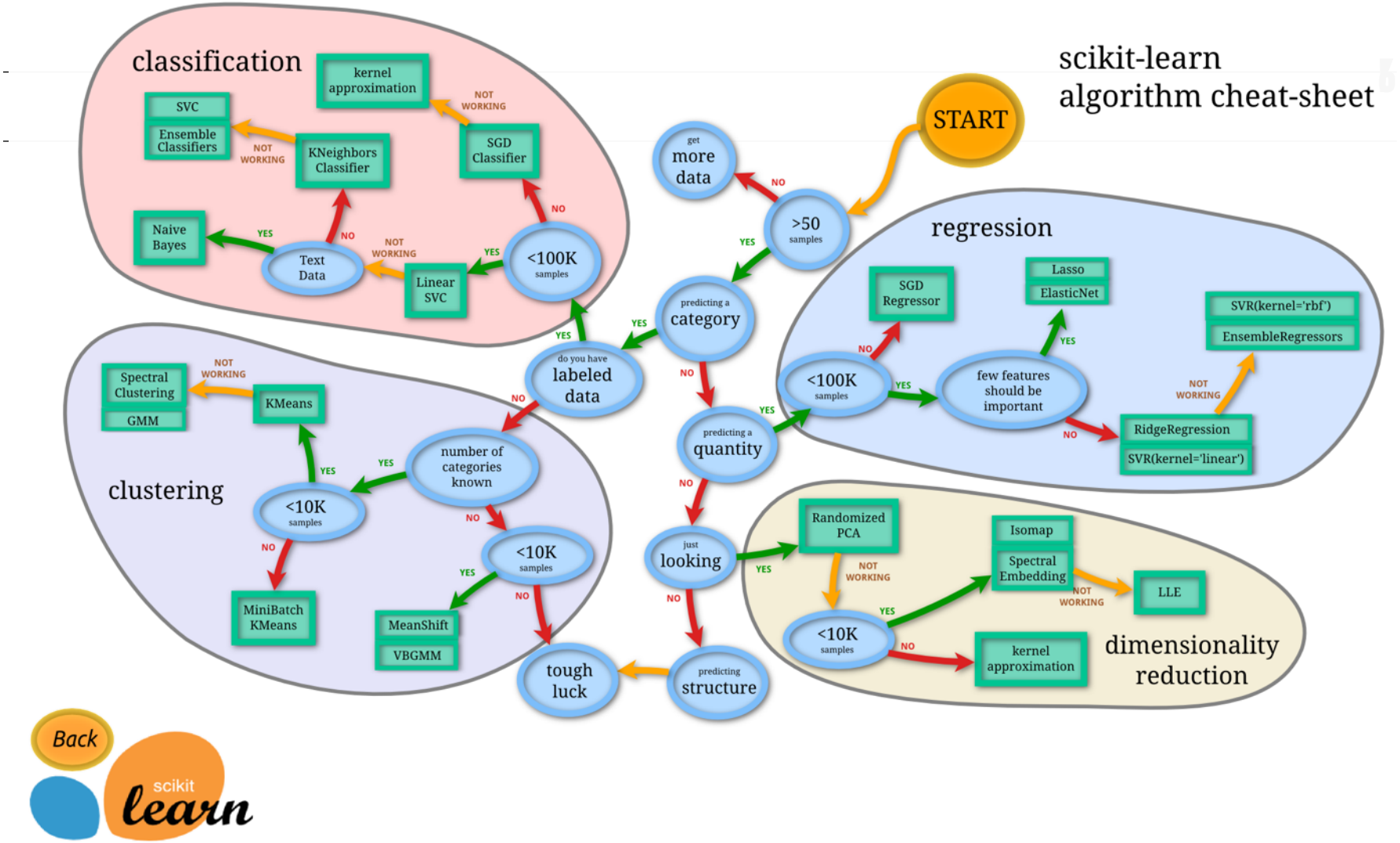
	continuous	***categorical***
supervised	*regression*	*classification*
unsupervised	*dimension reduction*	*clustering*

WE COVERED

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

WE COVERED

- *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*

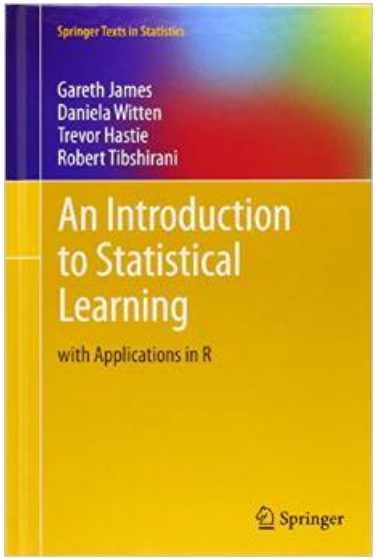

scikit-learn
algorithm cheat-sheet
START
get more data
NO
>50 samples
YES
predicting a category
YES
NO
do you have labeled data
YES
NO
predicting a quantity
YES
NO
just looking
YES
NO
predicting structure
tough luck
classification
kernel approximation
NOT WORKING
SVC
Ensemble Classifiers
NOT WORKING
KNeighbors Classifier
SGD Classifier
NO
Naive Bayes
YES
Text Data
NO
NOT WORKING
Linear SVC
YES
<100K samples
YES
regression
SGD Regressor
NO
<100K samples
YES
few features should be important
YES
Lasso
ElasticNet
NO
SVR(kernel='rbf')
EnsembleRegressors
NOT WORKING
RidgeRegression
SVR(kernel='linear')
clustering
Spectral Clustering
GMM
NOT WORKING
KMeans
YES
number of categories known
YES
NO
<10K samples
NO
MiniBatch KMeans
<10K samples
YES
NO
MeanShift
VBGMM
dimensionality reduction
Randomized PCA
YES
NOT WORKING
<10K samples
YES
NO
Isomap
Spectral Embedding
NOT WORKING
LLE
kernel approximation
Back
scikit
learn

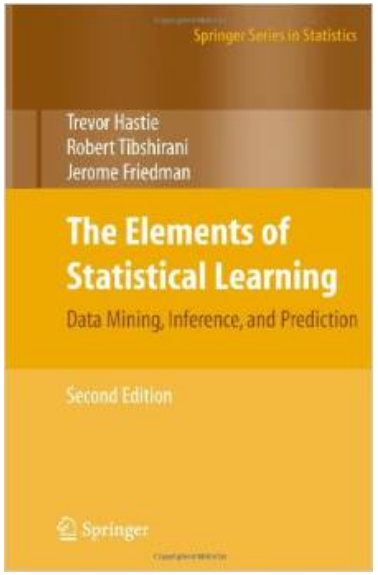
WHAT'S DONE

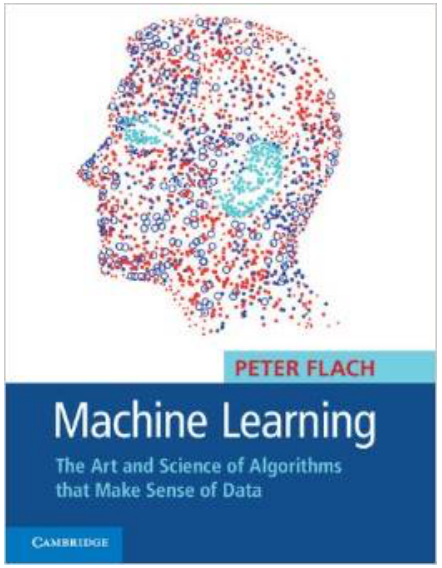
- *Python and iPython Notebook*
- *Scikit-learn and Pandas*
- *matplotlib and Seaborn*
- *Nltk and NetworkX*
- *SQL and SQLite*
- *Python recsys*

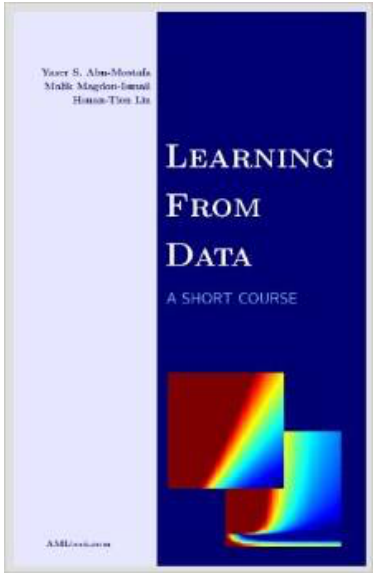
II. WHERE TO GO NEXT?

WHAT TO DO NEXT: READING

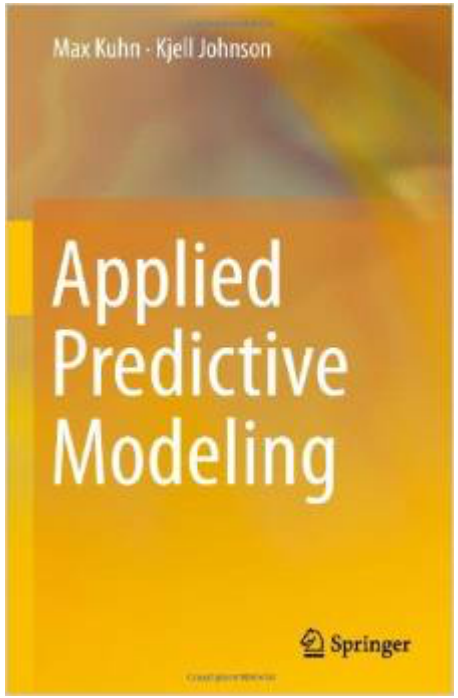


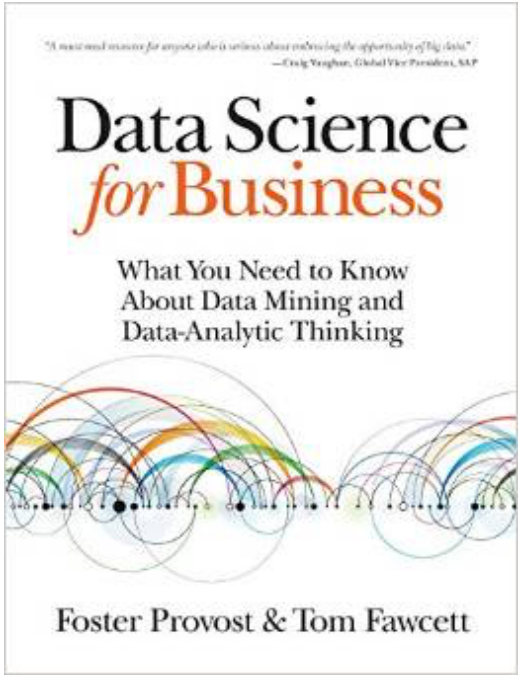


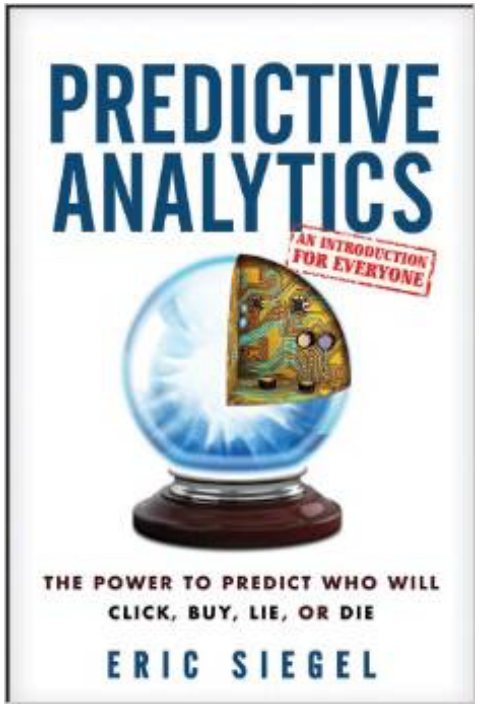




WHAT TO DO NEXT: READING





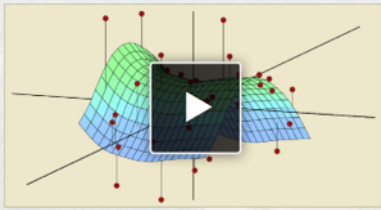


WHAT TO DO NEXT: READING

WHAT TO DO NEXT: ONLINE COURSES

Statistical Learning

REGISTER FOR STATLEARNING



Stanford University

Machine Learning

Ended 3 years ago

Course Record

Stanford University

Mining Massive Datasets

Ended 2 months ago

Course Record

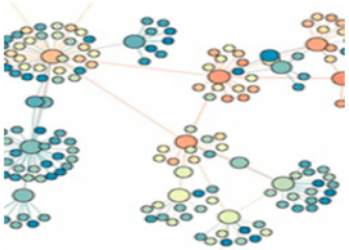
WHAT TO DO NEXT: ONLINE COURSES

Stanford University

Natural Language Processing

Ended 3 years ago

Course Record



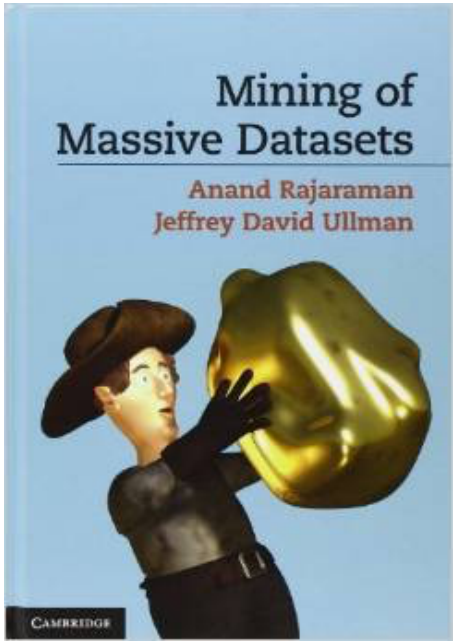
University of Michigan

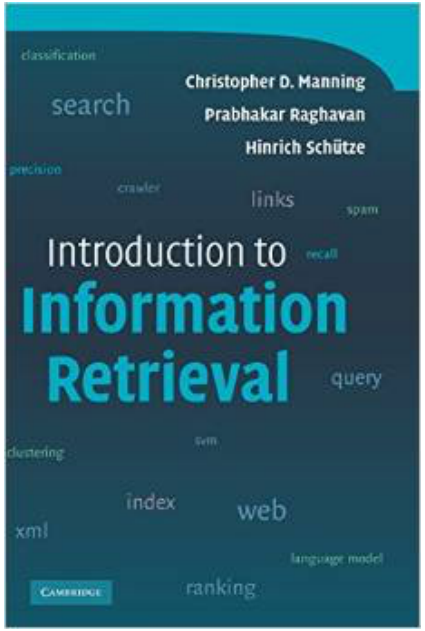
Social Network Analysis

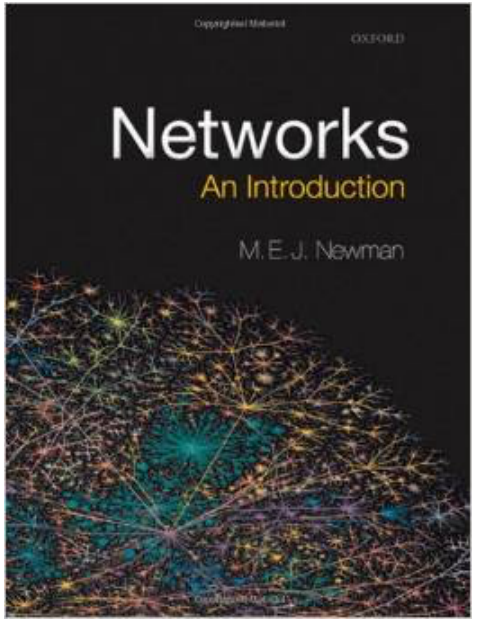
Ended a year ago

Course Record

WHAT TO DO NEXT: READING







WHAT TO DO NEXT: 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*

WHAT TO DO NEXT: COMPETITIONS AND PROJECTS

kaggle

[topcoder]™

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*

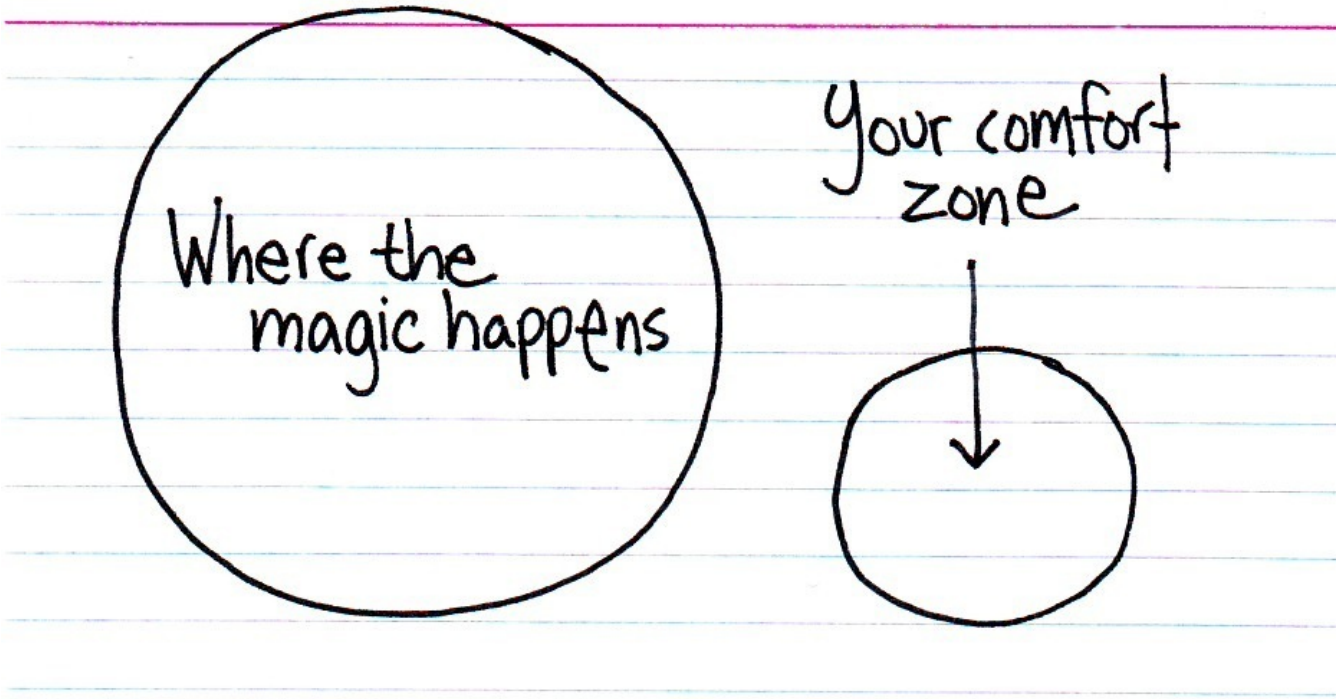
AND MOST OF ALL...

Each Other!!!

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

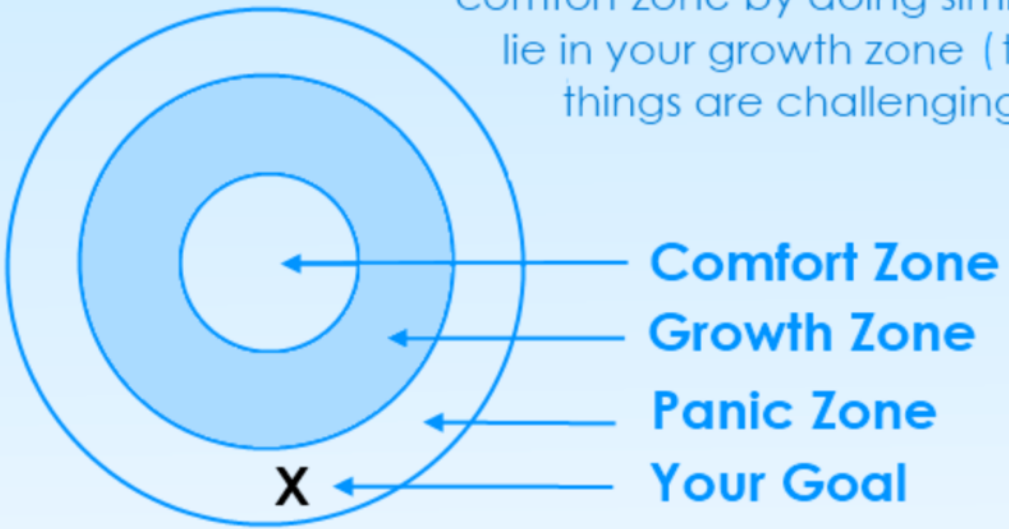
CONGRATULATIONS...

...for getting out of your comfort zone!



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.