

Machine Learning

Lecture 8, CMSC 170

John Roy Daradal / Instructor

Today's Topics

- Machine Learning
- ML Applications
- Datasets
- Evaluation
- ML Algorithms

**What do these things
have in common?**

AlphaGo



[<http://www.aiyellowpage.com/content/wp-content/uploads/2016/03/google-alphago-logo-540x334.png>]

Amazon Go



[<https://www.theverge.com/2016/12/5/13842592/amazon-go-new-cashier-less-convenience-store>]

Autocomplete



[<https://ulistentome.files.wordpress.com/2015/10/google-update-restrict-access-to-autocomplete-api-10th-august-2015.jpg>,

http://i.dailymail.co.uk/i/pix/2014/06/27/1403873221362_Image_galleryImage_THE_MOST_EMBARRASSING_SMA.JPG]

Facebook Face Recognition

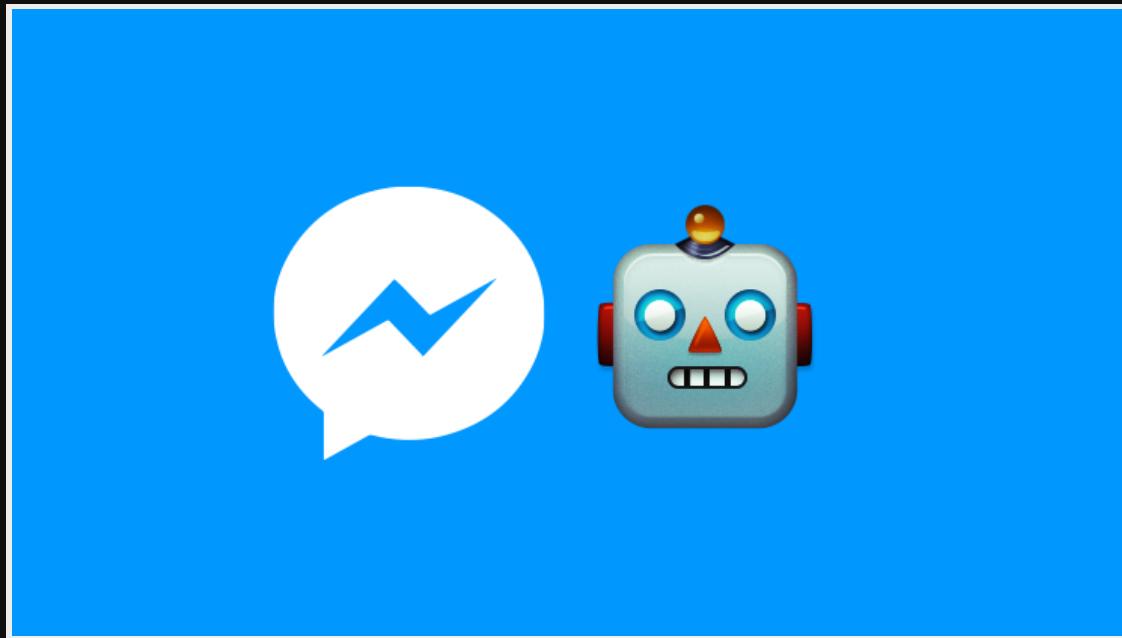
The photos you uploaded were grouped automatically so you can quickly label and notify friends in these pictures.
(Friends can always untag themselves.)

Who is this? Who is this? Who is this?

Who is this? Who is this? Who is this?

[<https://www.cnet.com/how-to/how-to-disable-facial-recognition-in-facebook/>]

Facebook Chat Bot



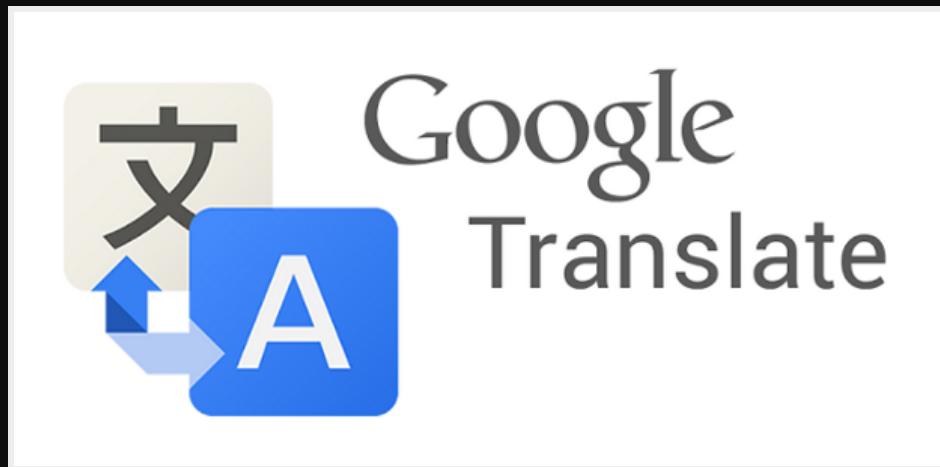
[<https://blog.hartleybrody.com/wp-content/uploads/2016/06/facebook-chatbot.png>]

Gmail Spam Filter



[<https://www.techlicious.com/images/computers/spam-filter-inboxes-email-shutterstock-510px.jpg>]

Google Translate



[<http://cdn.wccftech.com/wp-content/uploads/2015/01/google-translate-logo.png>]

Google Image Search

Google  JPG × game of thrones jon snow   

All **Images** Maps More Settings Tools

About 25,270,000,000 results (1.28 seconds)

 Image size:
4057 × 2700
Find other sizes of this image:
All sizes - Large

Best guess for this image: [game of thrones jon snow](#)

Jon Snow | Game of Thrones Wiki | FANDOM powered by Wikia
gameofthrones.wikia.com/wiki/Jon_Snow ▾
Jon Snow - born Aegon Targaryen - is the son of Lyanna Stark and Rhaegar Targaryen, the late...

Jon Snow and Daenerys Targaryen - Game of Thrones Wiki - Fandom
gameofthrones.wikia.com/wiki/Jon_Snow_and_Daenerys_Targaryen ▾
Jon Snow and Daenerys Targaryen first meet in the seventh season of Game ...

Visually similar images



Jon Snow 
Fictional character 

Ned Stark's bastard son, Jon joined the Night's Watch. On a mission for Lord Commander Mormont, Jon infiltrated the wildlings by pretending to forsake his Night Watch brothers. In doing so, he fell in love with Ygritte, a wildling woman. HBO

Portrayed by: [Kit Harington](#)

Title: [Lord Commander of the Night's Watch](#)

Parents: [Rhaegar Targaryen](#), [Lyanna Stark](#)

Significant others: [Ygritte](#), [Daenerys Targaryen](#) (TV series)

Siblings: [Sansa Stark](#), [Arya Stark](#), [Robb Stark](#), [Bran Stark](#), [Rickon Stark](#)

Notable aliases: [Lord Snow](#), [The Bastard of Winterfell](#), [King Crow](#), [The White Wolf](#), [Aegon Targaryen](#) (birth name, TV series)

People also search for

 Daenerys  Ygritte  Eddard  Tyrion  Arva Stark [View 15+ more](#)

[<https://www.google.com.ph/imghp?>]

Google Ads



[<https://edge45.co.uk/blog/google-adwords-evolution-timeline/>]

Google Self-Driving Cars



[<https://www.cnet.com/news/googles-self-driving-cars-still-need-that-human-drivers-touch/>]

Google Deep Dream



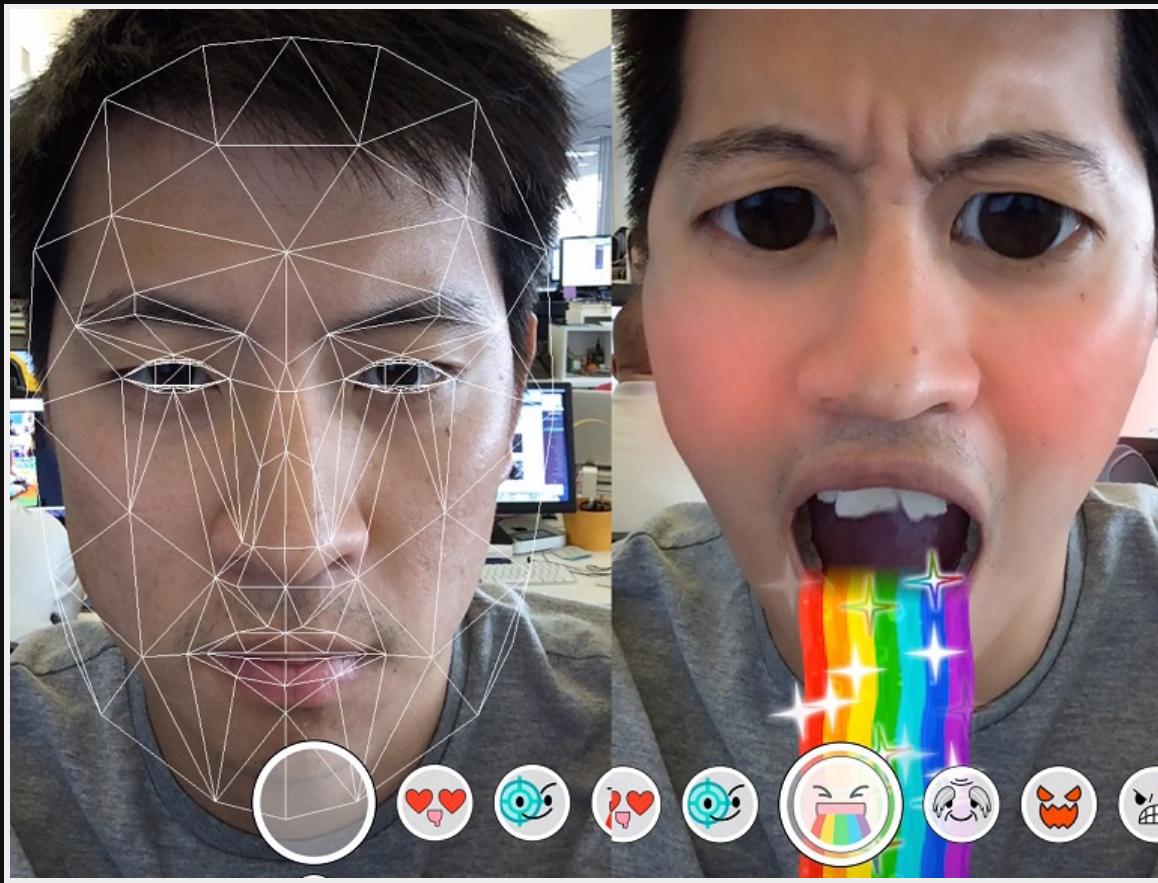
[<https://artoferickuns.files.wordpress.com/2015/07/dbb70e92b3b9dd88e1a8af2c4d0c068.jpg>]

Siri



[<http://ajournalofmusicalthings.com/wp-content/uploads/2017/07/Siri-app.jpg>]

Snapchat Filters



[<http://static1.businessinsider.com/image/55f97e93dd08952e3a8b4724/snapchats-new-selfie-filters-are-super-trippy.jpg>]

Waze



[<https://drjpau.files.wordpress.com/2015/06/waze.jpg?w=620>]

Yelp

Himawari 1721 reviews

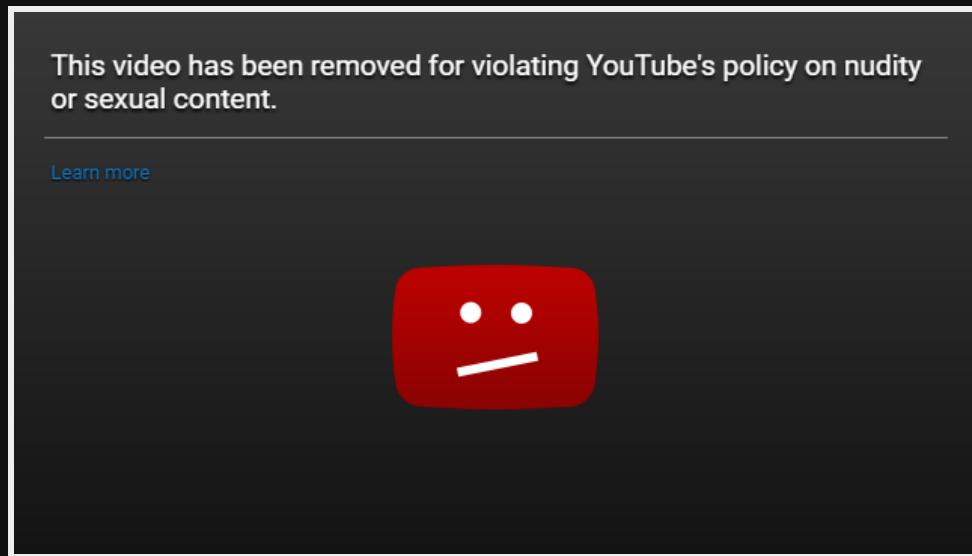
Add photos

All (1178) Food (1035) Interior (28) Outside (7) Drink (12) Menu (22)

The image shows a screenshot of a Yelp page for a restaurant named 'Himawari'. At the top, there's a yellow logo with the word 'Yelp'. Below it is the restaurant's name 'Himawari' with a small profile picture, a 5-star rating, and 1721 reviews. There's a red button to 'Add photos'. Below the header are navigation links: 'All (1178)', 'Food (1035)', 'Interior (28)' (which is underlined), 'Outside (7)', 'Drink (12)', and 'Menu (22)'. The main content is a grid of 28 photos. The first row contains six photos: 'Inside the restaurant', 'Love records on the walls!', 'Friday night', 'The interior', 'Seating inside', and 'The interior view'. The second row contains six photos: 'Always packed! 8:30 pm on a Friday', 'Kitchen', 'Interior', a photo of a menu board with text like 'Times opened and closed and a no outside...', 'Interior', and another interior shot. The third row contains five photos: 'Interior', 'Interior', 'Interior', a photo of a star-shaped light fixture, and 'Interior'. The fourth row contains three photos: 'Interior', 'Interior', and 'Interior'.

[<http://www.wordstream.com/images/machine-learning-yelp-image-categorizing.png>]

Youtube Content Detection



[https://cdn-images-1.medium.com/max/640/1*_5WvJBmq2Jlk9r0HX91vw.png]

YT Recommended Videos

YouTube PH

Search

Home

Trending

Subscriptions

LIBRARY

History

Watch later

2016

Deep Learning F...

Show more

SUBSCRIPTIONS

NBA 11

The Normies 10

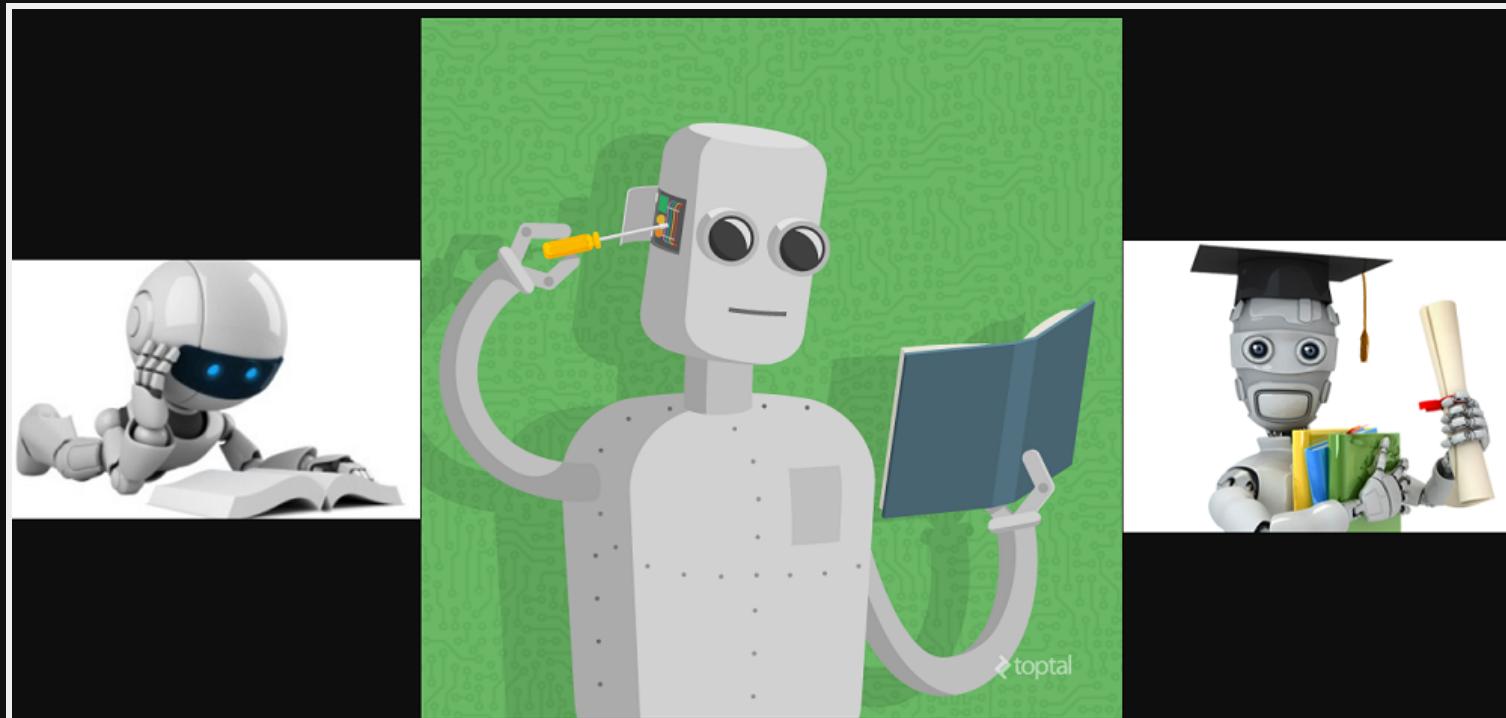
Recommended

 MY PREDICTED LIFE! 13:59 Wil Dasovich	 WAY BACK WHEN WITH DUA LIPA 6:16 ANDPOP	 Wil's 26th Birthday! 7:18 Charry Dasovich	 10 SURPRISING GAME OF THRONES REAL LIFE ACCENTS 12:58 Beanyman
MY SURVIVAL RATE REVEALED (Judgement Day) 168K views • 5 hours ago	Dua Lipa Talks Always Being The New Girl 75K views • 8 months ago	Wil's 26th birthday 21K views • 5 hours ago	10 'Game Of Thrones' Actors Whose Real-Life Accents 114K views • 3 weeks ago
 3:26 THE CHALLENGES WE FACE. Daniel Marsh	 1:41 Game of Thrones: Map Quiz // SiriusXM // Entertainment SiriusXM	 45:50 Maisie Williams - Pranks Fan - Funny Moments - Funny Mad Garage Fails	 3:00 BLACKPINK "Partition (Beyoncé) Dance Cover" • Fomo Daily
119K views • 2 days ago	84K views • 1 month ago	21K views • 3 weeks ago	4.1K views • 4 hours ago

[<https://www.youtube.com/>]

**What do these things
have in common?**

Machine Learning

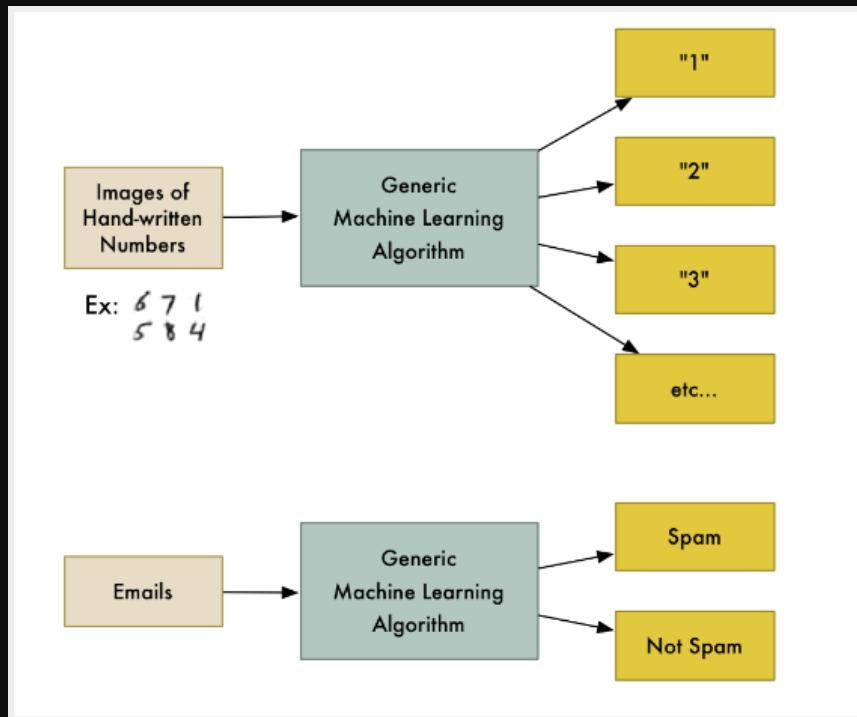


Machine Learning

*"Machine Learning gives computers the ability to **learn** without being **explicitly programmed**"*

- Arthur Samuel, 1959

Machine Learning



Machine Learning

*"A computer program **learns** from **experience E** with respect to **task T** and some **performance measure P**, if its performance on T , as measured by P , **improves** with experience E "*

- Tom Mitchell, 1997

Applications

Computer Vision



[https://www.cc.gatech.edu/~hays/compvision/comp_vision_teaser_by_kirkh.jpg]

Facebook Face Recognition

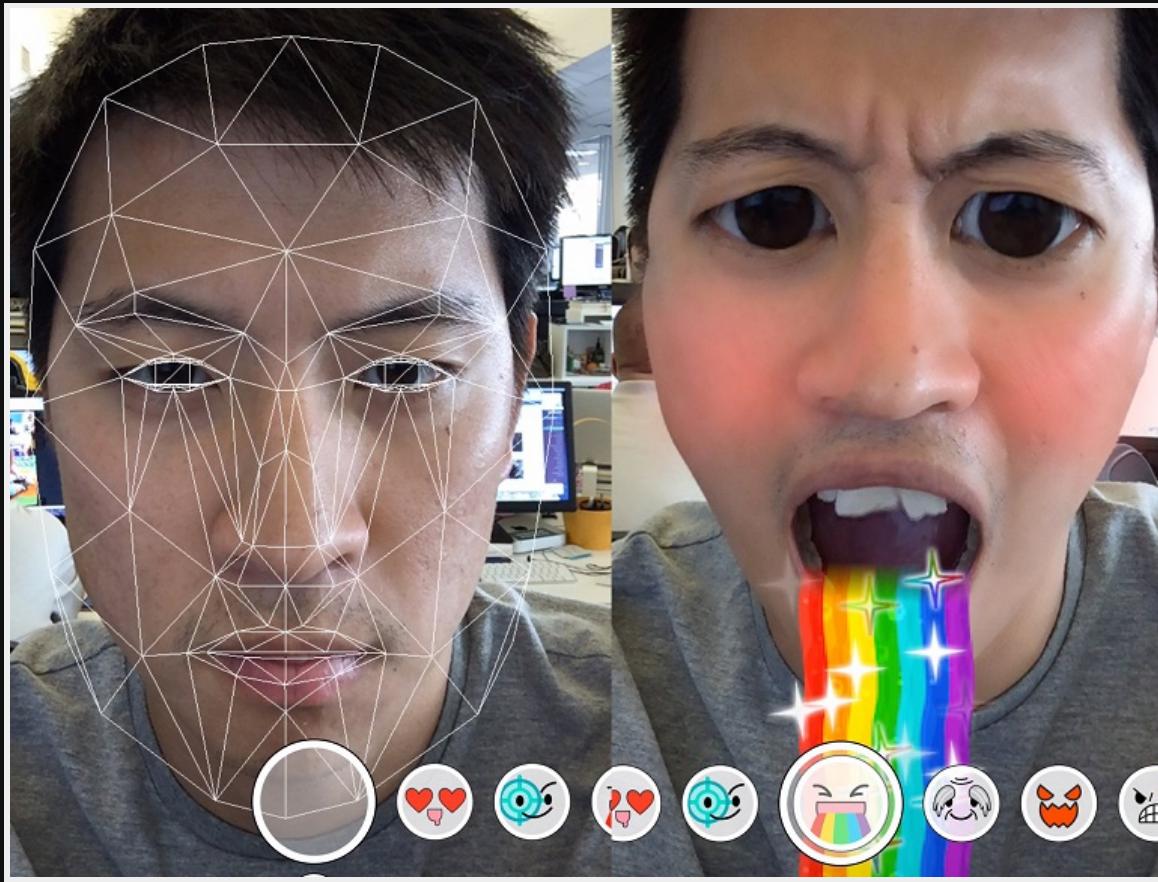
The photos you uploaded were grouped automatically so you can quickly label and notify friends in these pictures.
(Friends can always untag themselves.)

Who is this? Who is this? Who is this?

Who is this? Who is this? Who is this?

[<https://www.cnet.com/how-to/how-to-disable-facial-recognition-in-facebook/>]

Snapchat Filters



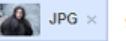
[<http://static1.businessinsider.com/image/55f97e93dd08952e3a8b4724/snapchats-new-selfie-filters-are-super-trippy.jpg>]

Amazon Go



[<https://www.theverge.com/2016/12/5/13842592/amazon-go-new-cashier-less-convenience-store>]

Google Image Search

Google  game of thrones jon snow   

All **Images** Maps More Settings Tools

About 25,270,000,000 results (1.28 seconds)

 Image size:
4057 x 2700
Find other sizes of this image:
All sizes - Large

Best guess for this image: [game of thrones jon snow](#)

Jon Snow | Game of Thrones Wiki | FANDOM powered by Wikia
gameofthrones.wikia.com/wiki/Jon_Snow ▾
Jon Snow - born Aegon Targaryen - is the son of Lyanna Stark and Rhaegar Targaryen, the late...

Jon Snow and Daenerys Targaryen - Game of Thrones Wiki - Fandom
gameofthrones.wikia.com/wiki/Jon_Snow_and_Daenerys_Targaryen ▾
Jon Snow and Daenerys Targaryen first meet in the seventh season of Game ...

Visually similar images



Jon Snow 
Fictional character 

Ned Stark's bastard son, Jon joined the Night's Watch. On a mission for Lord Commander Mormont, Jon infiltrated the wildlings by pretending to forsake his Night Watch brothers. In doing so, he fell in love with Ygritte, a wildling woman. HBO

Portrayed by: Kit Harington
Title: Lord Commander of the Night's Watch
Parents: Rhaegar Targaryen, Lyanna Stark
Significant others: Ygritte, Daenerys Targaryen (TV series)
Siblings: Sansa Stark, Arya Stark, Robb Stark, Bran Stark, Rickon Stark
Notable aliases: Lord Snow, The Bastard of Winterfell, King Crow, The White Wolf, Aegon Targaryen (birth name, TV series)

People also search for 
[View 15+ more](#)

Google Self-Driving Cars



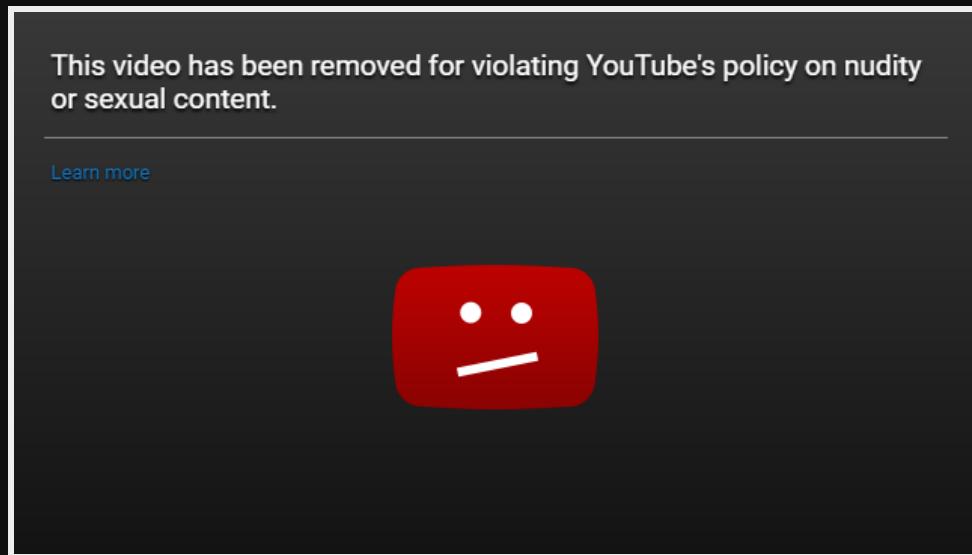
[<https://www.cnet.com/news/googles-self-driving-cars-still-need-that-human-drivers-touch/>]

Google Deep Dream



[<https://artoferickuns.files.wordpress.com/2015/07/dbb70e92b3b9dd88e1a8af2c4d0c068.jpg>]

Youtube Content Detection



[https://cdn-images-1.medium.com/max/640/1*_5WvJBmq2Jlk9r0HX91vw.png]

Yelp

Himawari 1721 reviews

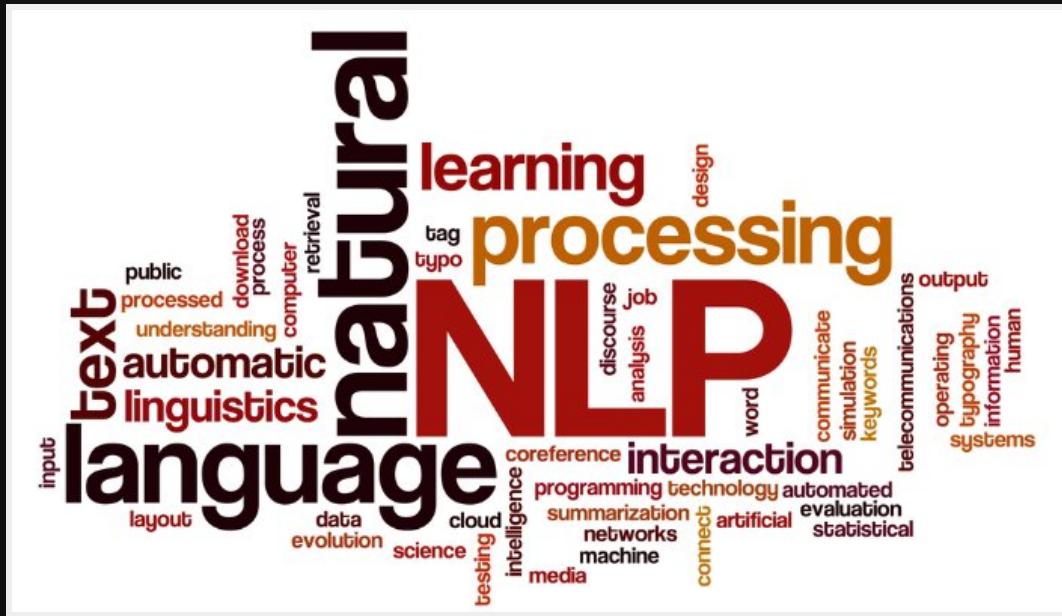
Add photos

All (1178) Food (1035) Interior (28) Outside (7) Drink (12) Menu (22)

The image shows a screenshot of a Yelp page for a restaurant named 'Himawari'. At the top, there's a yellow logo with the word 'Yelp'. Below it is the restaurant's name 'Himawari' with a small profile picture, a 5-star rating, and 1721 reviews. There's a red button to 'Add photos'. Below the header are navigation links: 'All (1178)', 'Food (1035)', 'Interior (28)' (which is underlined), 'Outside (7)', 'Drink (12)', and 'Menu (22)'. The main content is a grid of 28 small images. The first row contains six images: 'Inside the restaurant', 'Love records on the walls!', 'Friday night', 'The interior', 'Seating inside', and 'The interior view'. The second row contains six images: 'Always packed! 8:30 pm on a Friday', 'Kitchen', 'Interior', 'Times opened and closed and a no outside...', 'Interior', and another 'Interior' shot. The third row contains five images: 'Interior', 'Interior', 'Interior', 'Interior', and 'Interior'. Each image has a caption below it.

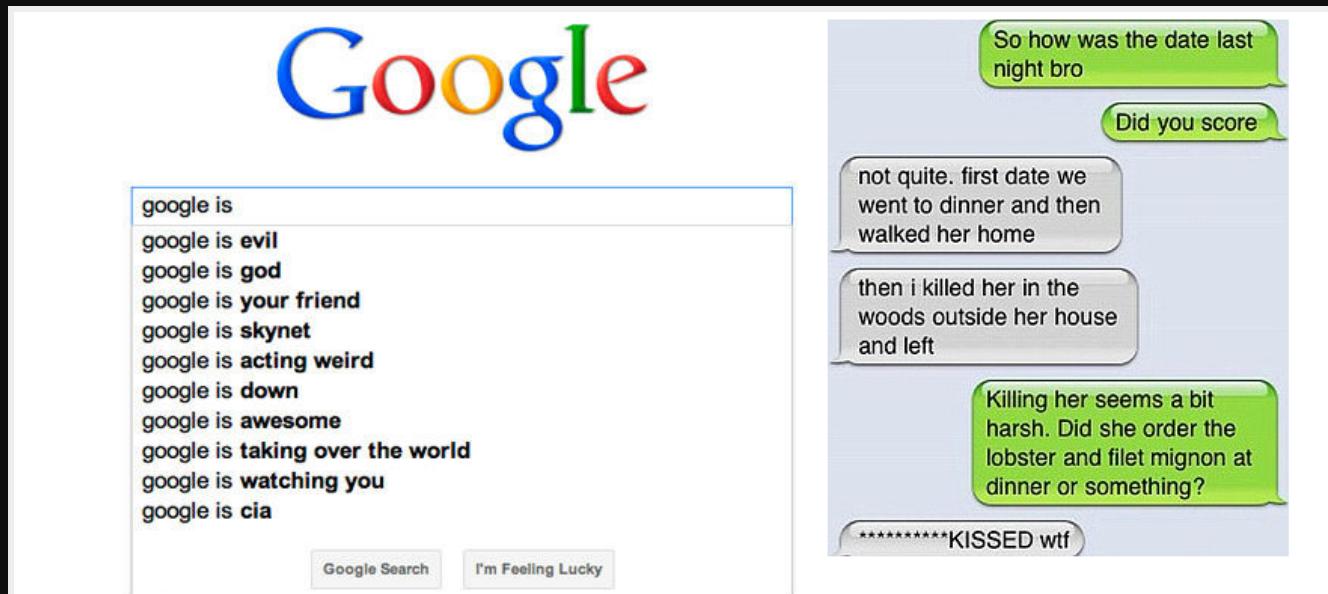
[<http://www.wordstream.com/images/machine-learning-yelp-image-categorizing.png>]

Natural Language Processing



[<https://media.licdn.com/mpr/mpr/AAEAAQAAAAAAAARfAAAJDY5Mzk2NjUwLWQ1MWEtNDhiYS1hNmNkLTRkYzc5NTNiOTQ1YQ.jpg>]

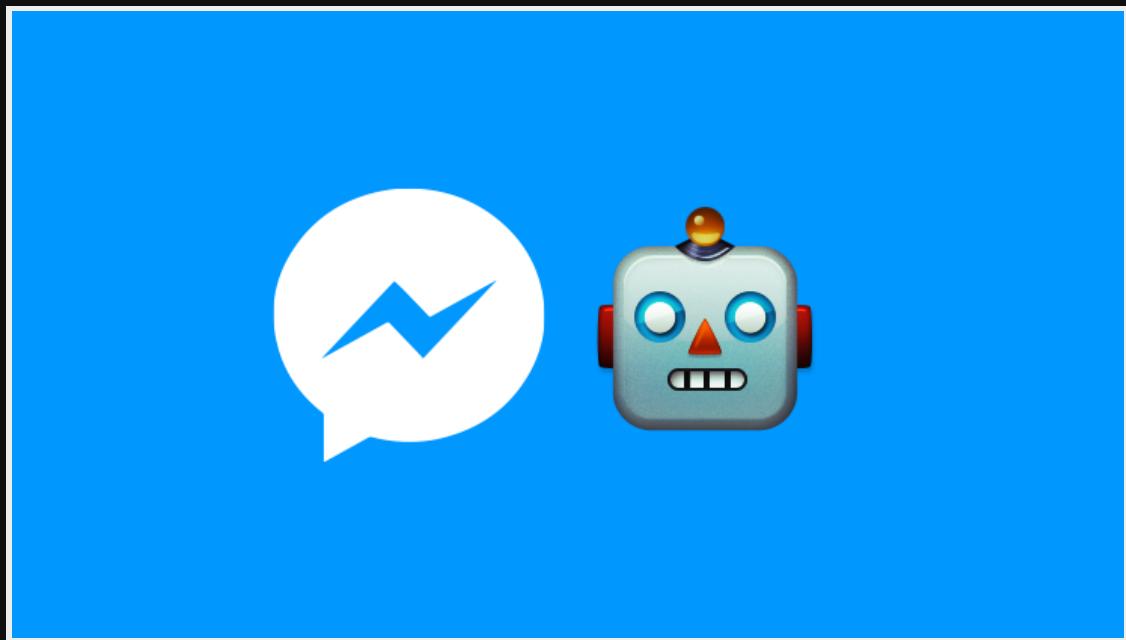
Autocomplete



[<https://ulistentome.files.wordpress.com/2015/10/google-update-restrict-access-to-autocomplete-api-10th-august-2015.jpg>,

http://i.dailymail.co.uk/i/pix/2014/06/27/1403873221362_Image_galleryImage_THE_MOST_EMBARRASSING_SMA.JPG]

Facebook Chat Bot



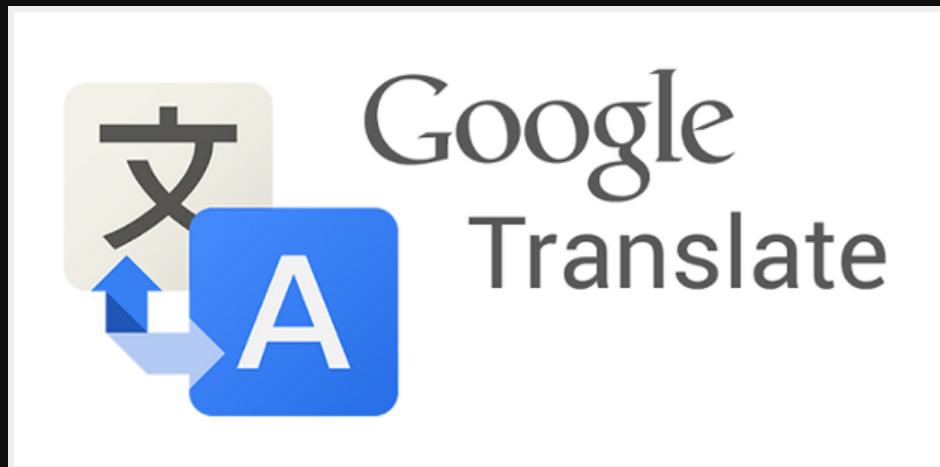
[<https://blog.hartleybrody.com/wp-content/uploads/2016/06/facebook-chatbot.png>]

Gmail Spam Filter



[<https://www.techlicious.com/images/computers/spam-filter-inboxes-email-shutterstock-510px.jpg>]

Google Translate



[<http://cdn.wccftech.com/wp-content/uploads/2015/01/google-translate-logo.png>]

Speech Recognition



[<http://copia.com.au/wp-content/themes/copia/images/professionalspeech.png>]

Siri



[<http://ajournalofmusicalthings.com/wp-content/uploads/2017/07/Siri-app.jpg>]

Recommender Systems



[<https://www.slideshare.net/xamat/qcon-sf-2013-machine-learning-recommender-systems-netflix-scale>]

YT Recommended Videos

YouTube PH

Search

Home

Trending

Subscriptions

LIBRARY

History

Watch later

2016

Deep Learning F...

Show more

SUBSCRIPTIONS

NBA 11

The Normies 10

Recommended

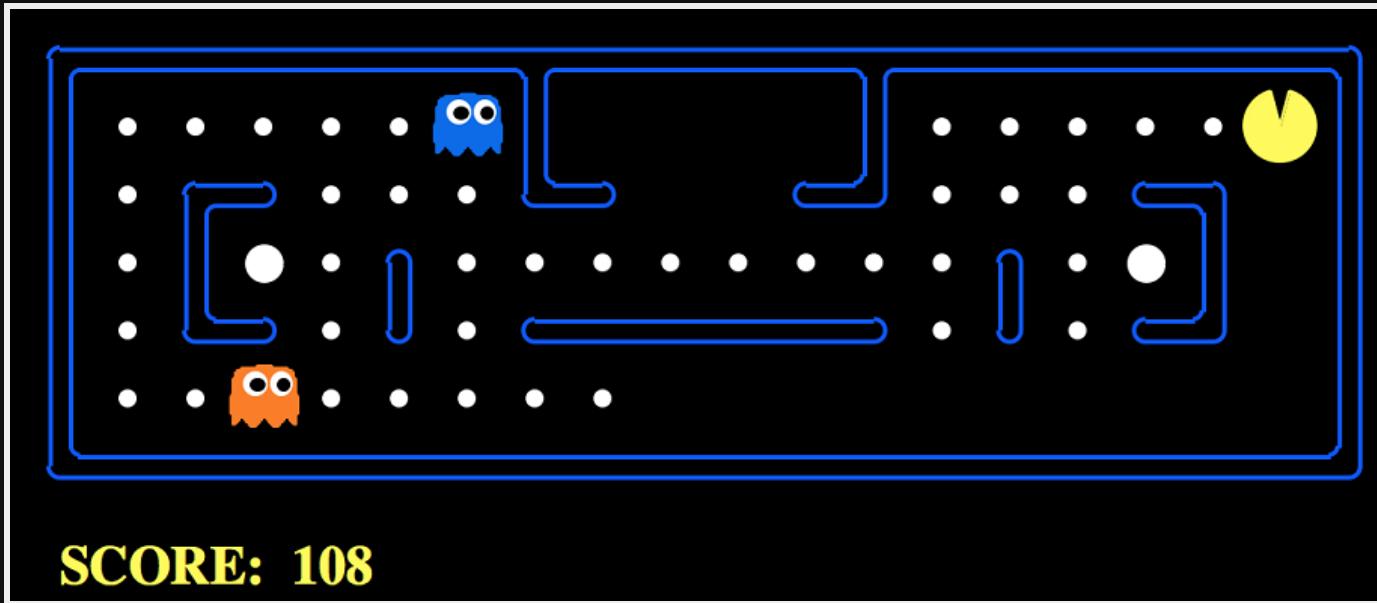
 MY PREDICTED LIFE! 13:59 Wil Dasovich	 WAY BACK WHEN WITH DUA LIPA 6:16 ANDPOP	 Wil's 26th Birthday! 7:18 Charry Dasovich	 10 SURPRISING GAME OF THRONES REAL LIFE ACCENTS 12:58 Beanyman
MY SURVIVAL RATE REVEALED (Judgement Day) 168K views • 5 hours ago	Dua Lipa Talks Always Being The New Girl 75K views • 8 months ago	Wil's 26th birthday 21K views • 5 hours ago	10 'Game Of Thrones' Actors Whose Real-Life Accents 114K views • 3 weeks ago
 3:26 THE CHALLENGES WE FACE. Daniel Marsh	 1:41 Game of Thrones: Map Quiz // SiriusXM // Entertainment SiriusXM	 45:50 Maisie Williams - Pranks Fan - Funny Moments - Funny Mad Garage Fails	 3:00 BLACKPINK "Partition (Beyoncé) Dance Cover" • Fomo Daily
119K views • 2 days ago	84K views • 1 month ago	21K views • 3 weeks ago	4.1K views • 4 hours ago

Google Ads



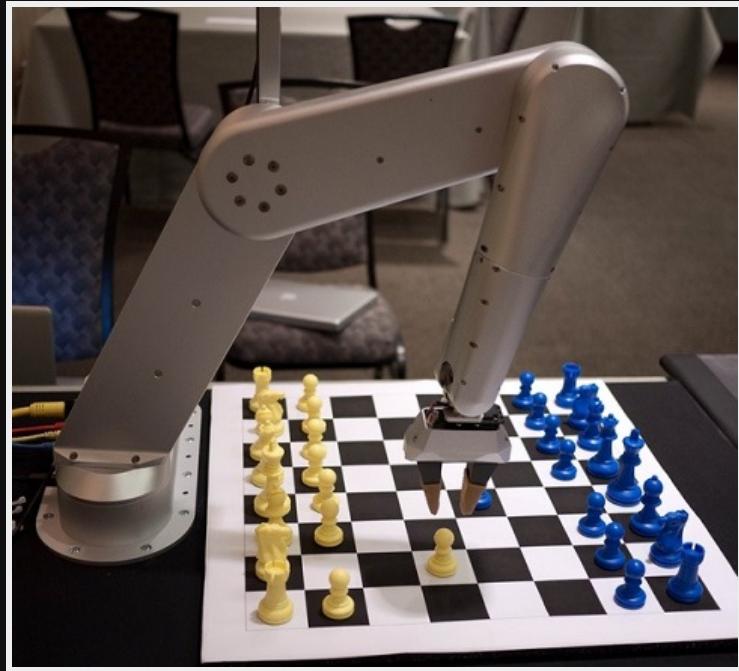
[<https://edge45.co.uk/blog/google-adwords-evolution-timeline/>]

Game Playing



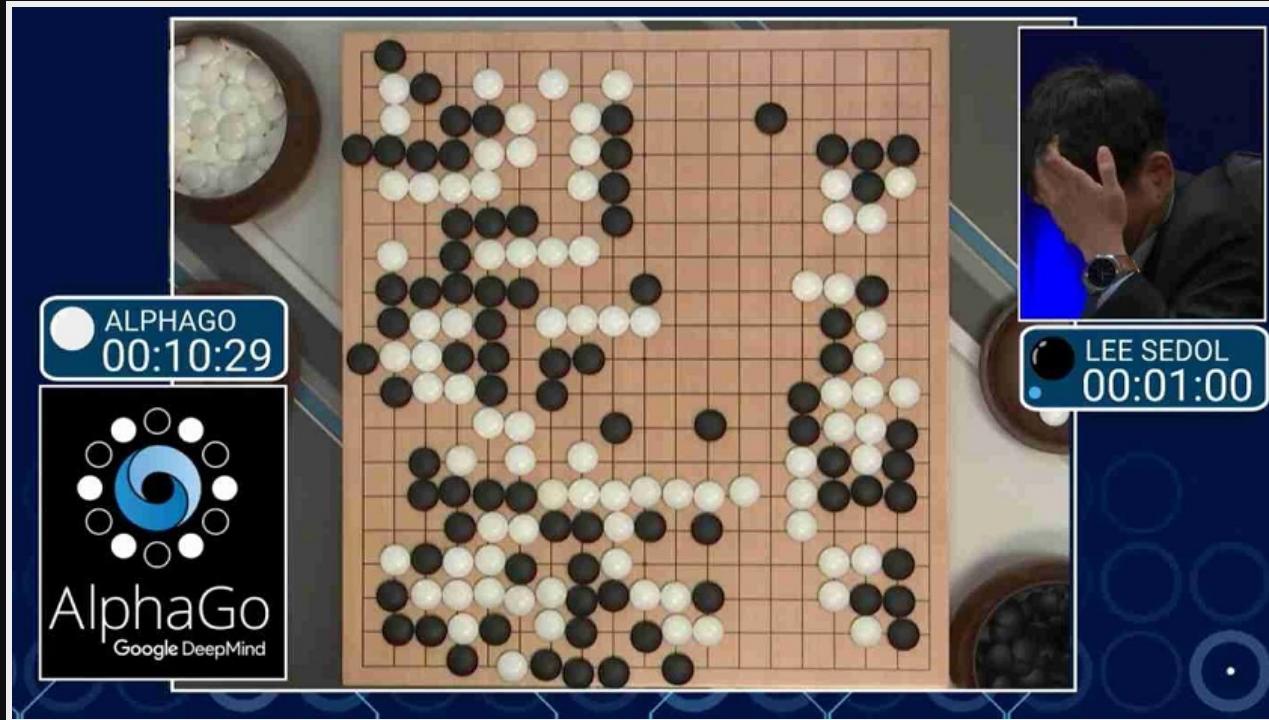
[<http://cs.wellesley.edu/~cs232/modularity/pacman.png>]

Chess



[<https://qph.ec.quoracdn.net/main-qimg-5ed0c1dce4bef40dbb37dcd15a0d1208-c>]

AlphaGo



[<https://gogameguru.com/i/2016/03/AlphaGo-Lee-Sedol-game-3-game-over.jpg>]

Machine Learning

Supervised

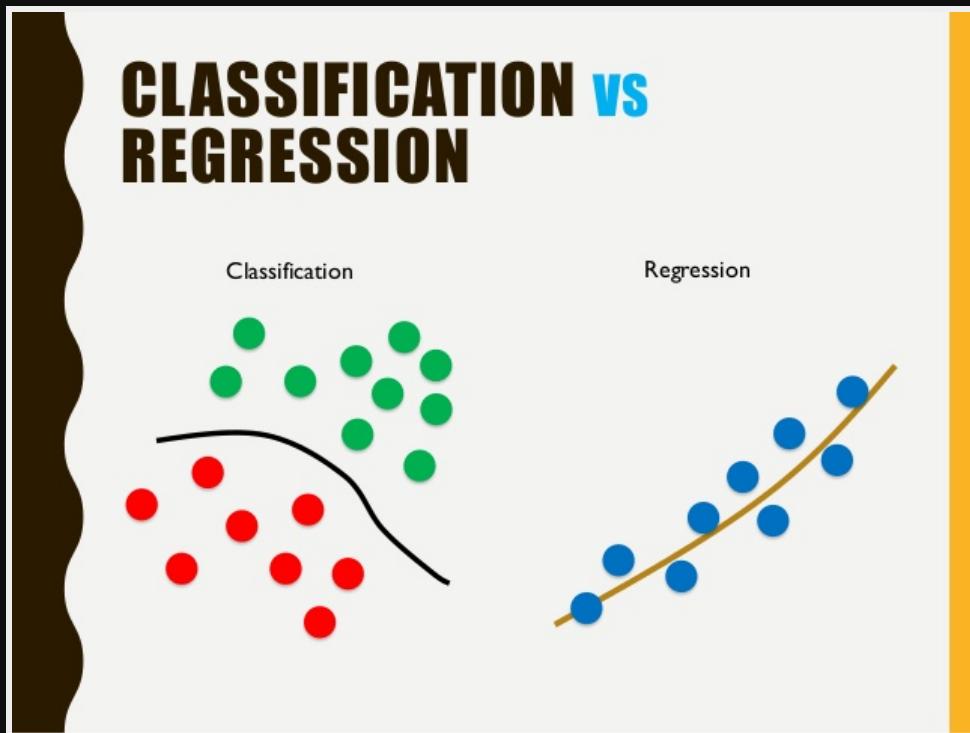
- *Input:* **labeled** data samples (x, y)
- x = data, y = label
- *Task:* **classification** or **regression**

Unsupervised

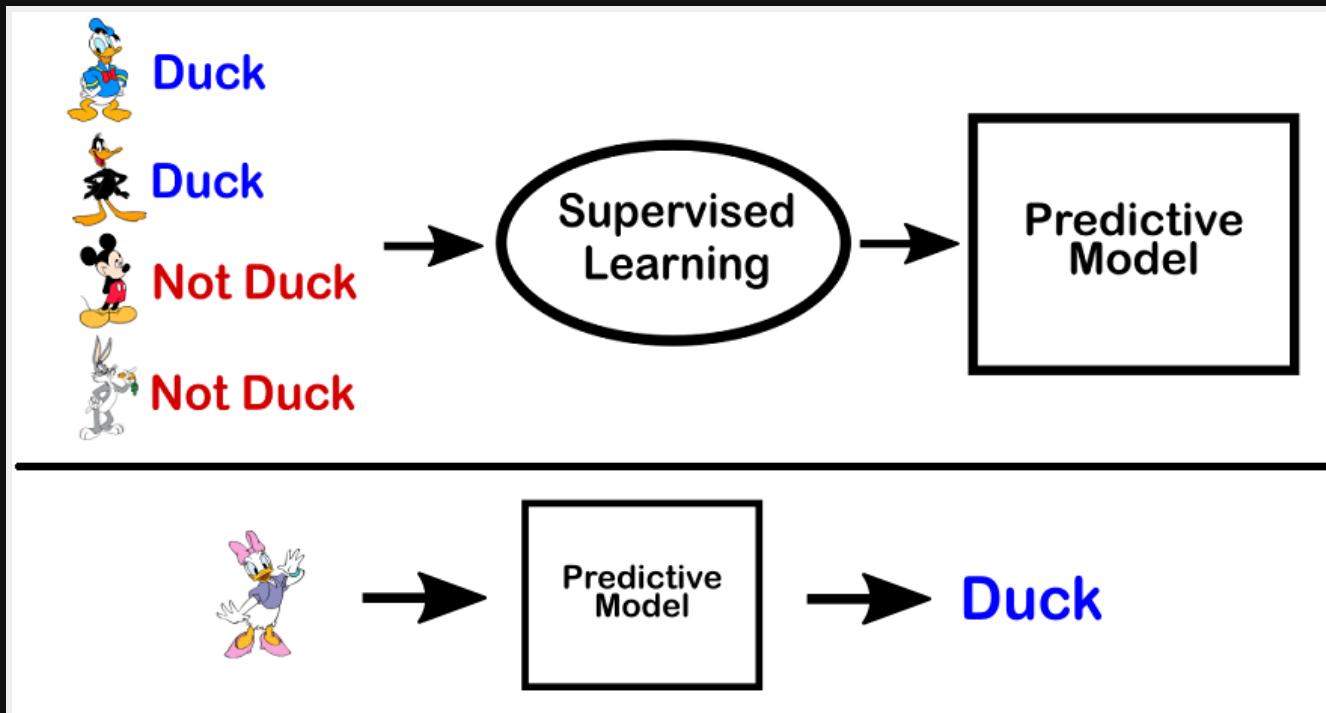
- *Input:* **unlabeled** data
- *Task:* find patterns, **clustering**, association

Also: **semi-supervised** (some labeled), **reinforcement** (games)

Supervised Learning



Supervised Learning



Supervised Learning

label = 5



label = 0



label = 4



label = 1



label = 9



label = 2



label = 1



label = 3



label = 1



label = 4



label = 3



label = 5



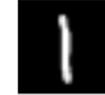
label = 3



label = 6



label = 1



label = 7



label = 2



label = 8



label = 6



label = 9

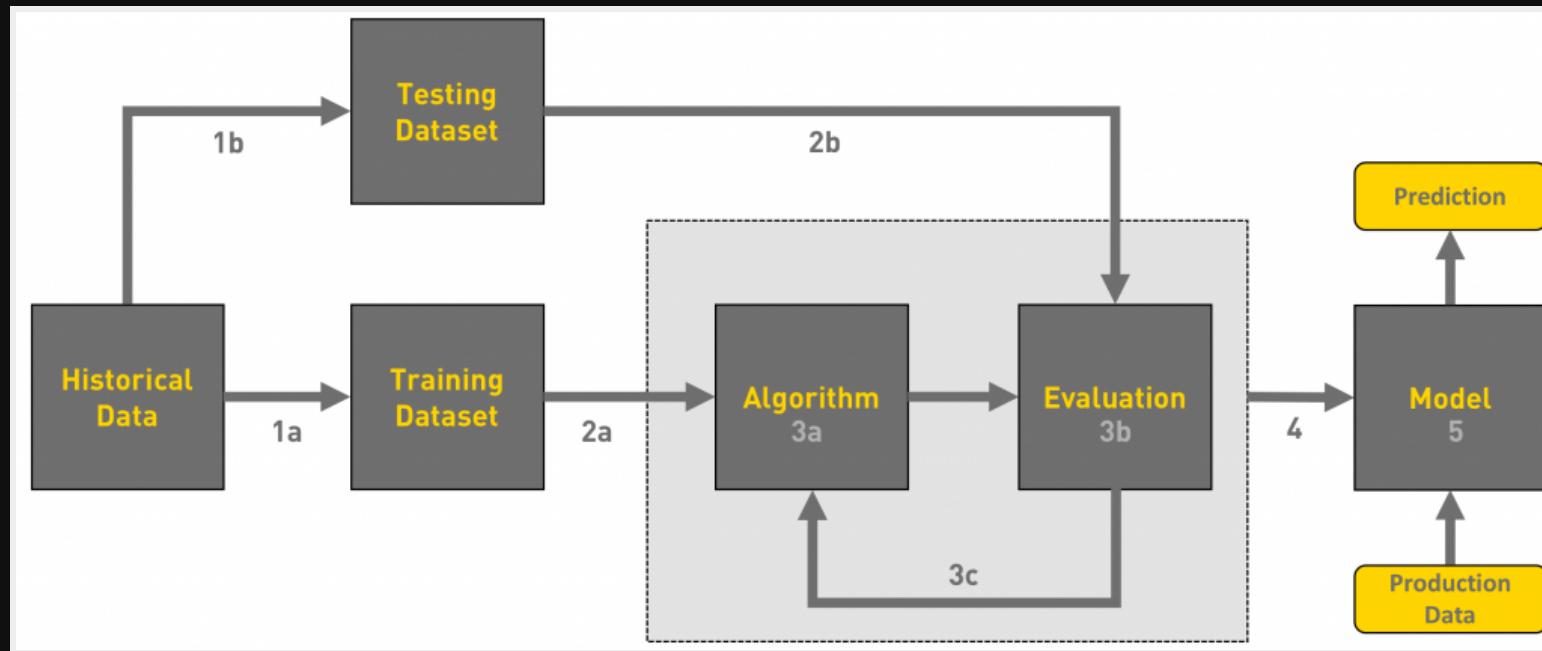


Supervised Learning

Analogy: Teaching a Child



Workflow



Ingredients

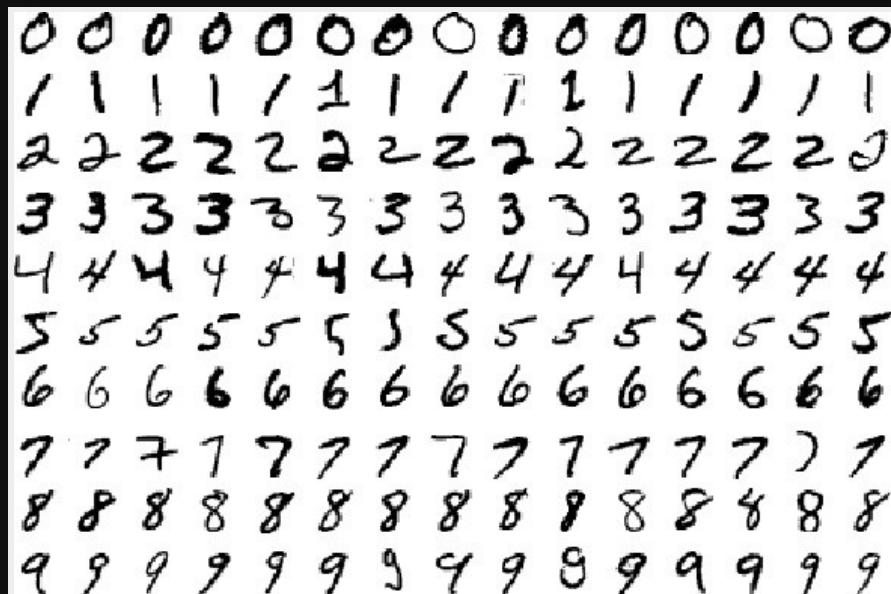
1. Data (*lots* of data)
2. Training Algorithm
3. Evaluation Metrics

Data

- **Data is king** in machine learning
(not *algorithms*)
- More data = *better* performance
- Data = images + labels (*supervised*)

MNIST Dataset

Handwritten digits (0-9), 70K samples



ImageNet Dataset

1K classes, ~1.3M images



Spam Dataset

- Enron Spam Dataset
- SMS Spam Dataset
- Kaggle Spam Dataset

Data

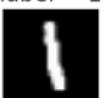
- *Divided* into: **training**, **test**, **validation** set
- Data **splits**: 70-15-15, 70-20-10, 80-10-10

Datasets

Training set

- used by *algorithm* in **training**
- classifier learns from these data samples (image + label)
- takes up bulk of dataset (~70-80%)

Example

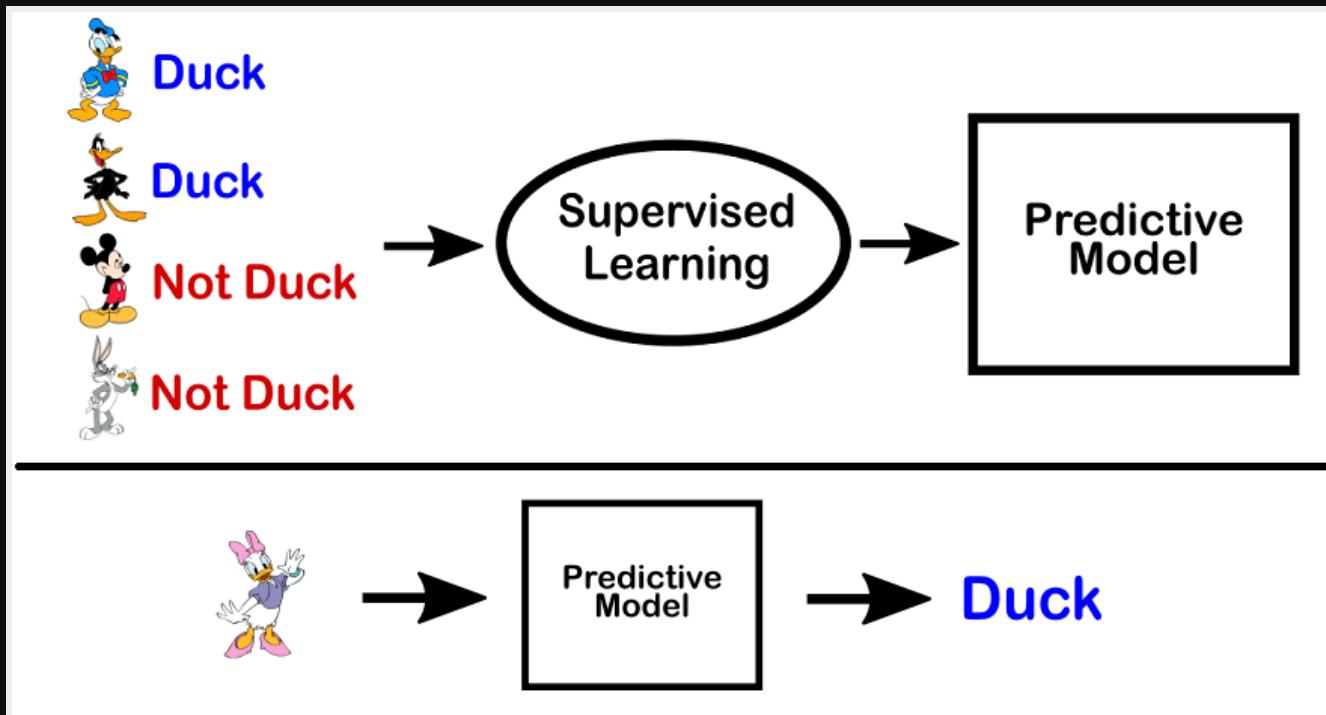
label = 5	label = 0	label = 4	label = 1	label = 9
				
label = 2	label = 1	label = 3	label = 1	label = 4
				
label = 3	label = 5	label = 3	label = 6	label = 1
				
label = 7	label = 2	label = 8	label = 6	label = 9
				

Datasets

Test set

- used to **evaluate performance** of classifier **after training**
- does classifier **generalize** well to images it hasn't seen before?

Example



Datasets

Validation set

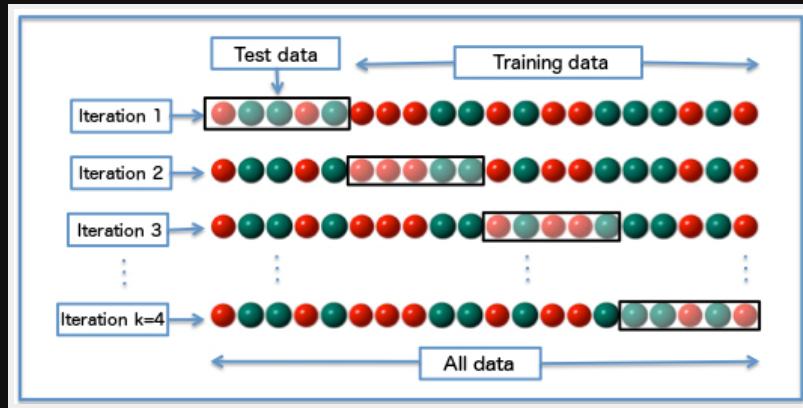
- *test set* used during **training**
- used to **estimate** classifier performance
- used to determine whether to *continue* or *stop* training

Datasets

Validation vs Test Set

- both **held-back** data (not in *training*)
- **validation** set: **biased** evaluation
(model "sees" it in training)
- **test** set: **unbiased** evaluation
(not seen in training)

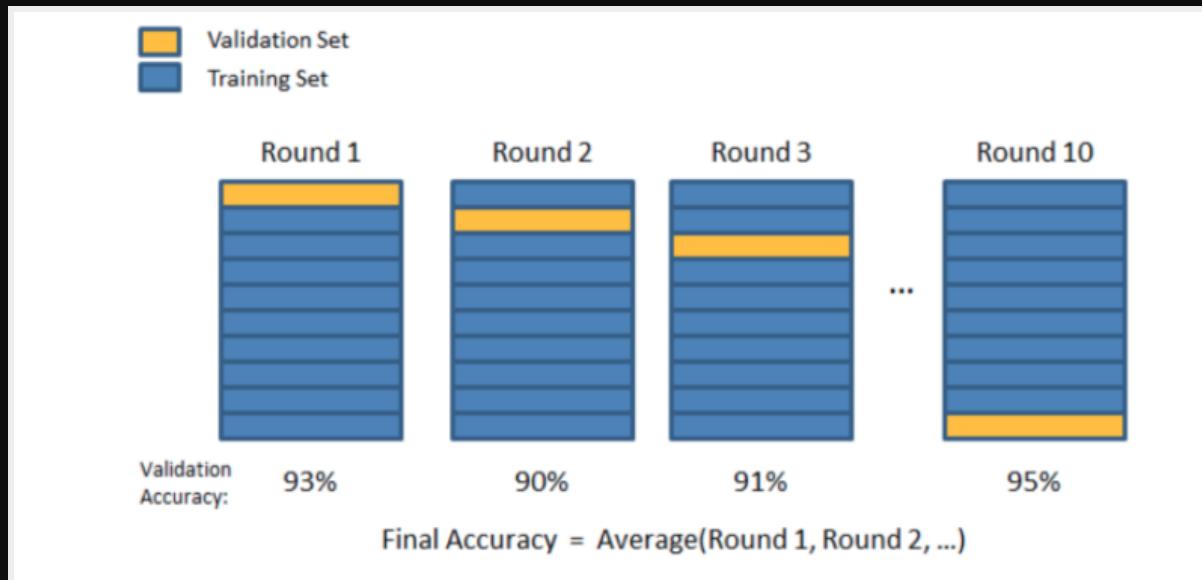
Cross-Validation



Cross-Validation

- Divide training set into **k folds**
- **Validation**: use *different fold* each iteration
- Used when **dataset is small**
- Computationally *expensive*

Cross-Validation



Evaluation

Testing **how good** the classifier is

```
for (data, trueLabel) in testSet:  
    predictedLabel = model.predict(data)  
    compare(trueLabel, predictedLabel)
```

Binary Classification

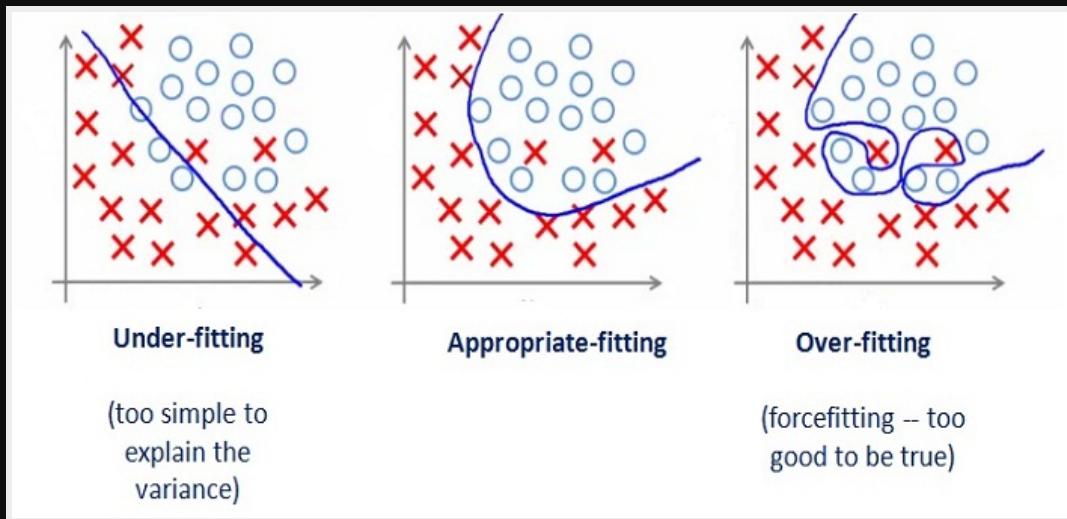
- **True Positive**: correct = 1, pred = 1
- **True Negative**: correct = 0, pred = 0
- **False Positive**: correct = 0, pred = 1
- **False Negative**: correct = 1, pred = 0

Performance

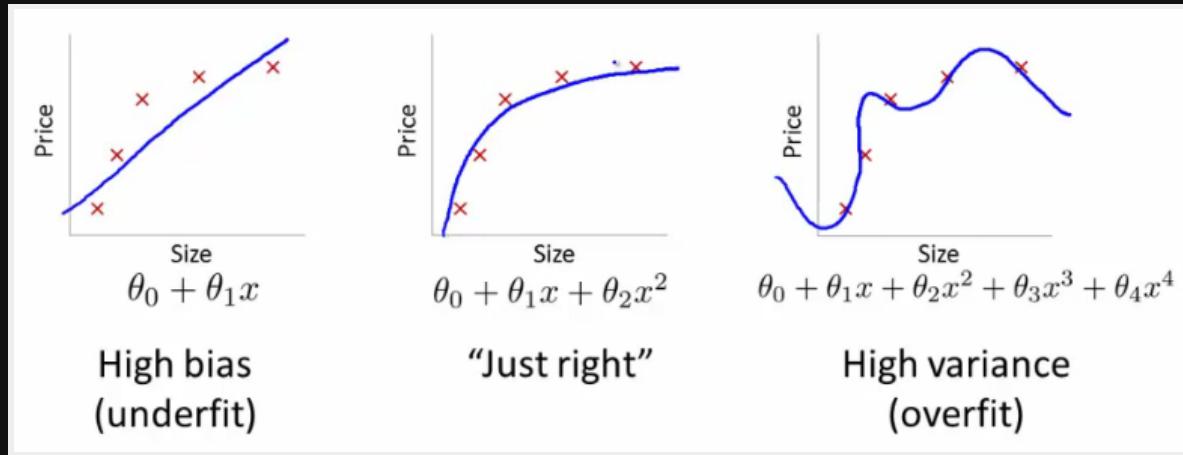
Training error

- measures classifier performance on *training data*
- not very important
- *low training error* = **good classifier** or **overfitted classifier**

Overfitting & Underfitting



Overfitting & Underfitting



Performance

Test error

- measures how well the classifier **generalizes** to *unseen data*
- the **most important** of the three
- *low test error = good classifier*

Performance

Validation error

- **estimates** *test error* during *training*
- helps in *adjusting model parameters* during training

Golden Rule

"Do not use test data in training"

Evaluation Metrics

Accuracy

- % of **correct answers**
- correct / total
- $(\text{TP}+\text{TN}) / (\text{TP}+\text{TN}+\text{FP}+\text{FN})$

Evaluation Metrics

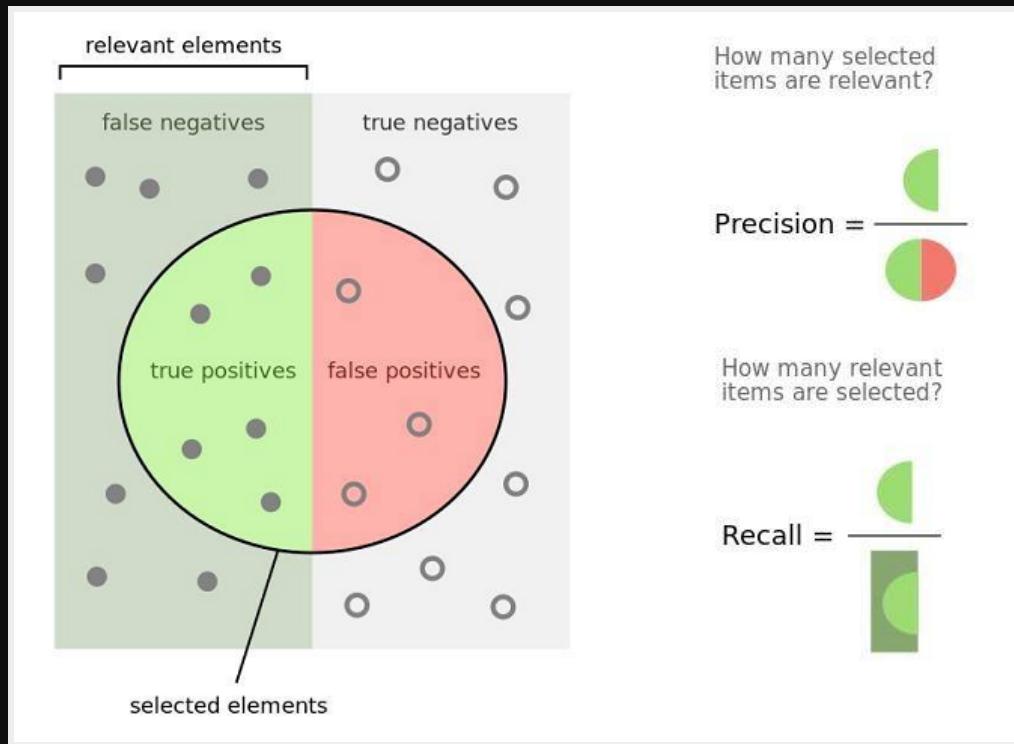
Precision

- % of **correct positive predictions**
- $\text{TP} / (\text{TP} + \text{FP})$

Recall

- % of **positive samples correctly classified as positive**
- $\text{TP} / (\text{TP} + \text{FN})$

Precision vs Recall



Example

- Dog vs. Cat classifier, *interest*: dog
- **Precision**: out of samples classified as dog, how many are actual dogs?
- **Recall**: out of actual dog samples, how many were classified as dogs?

Example

- Spam classifier
- **Precision**: out of all predicted spam, how many are actual spam?
- **Recall**: out of all actual spam, how many were predicted as spam?

Precision & Recall

high precision, high recall

- very good

low precision, low recall

- very bad

Precision & Recall

low precision, high recall

- **aggressive**: returns many positive answer, most are wrong
- "more entries you send, more chances of winning"
- *Example*: mostly predicts spam; lots of important emails end up in spam folder

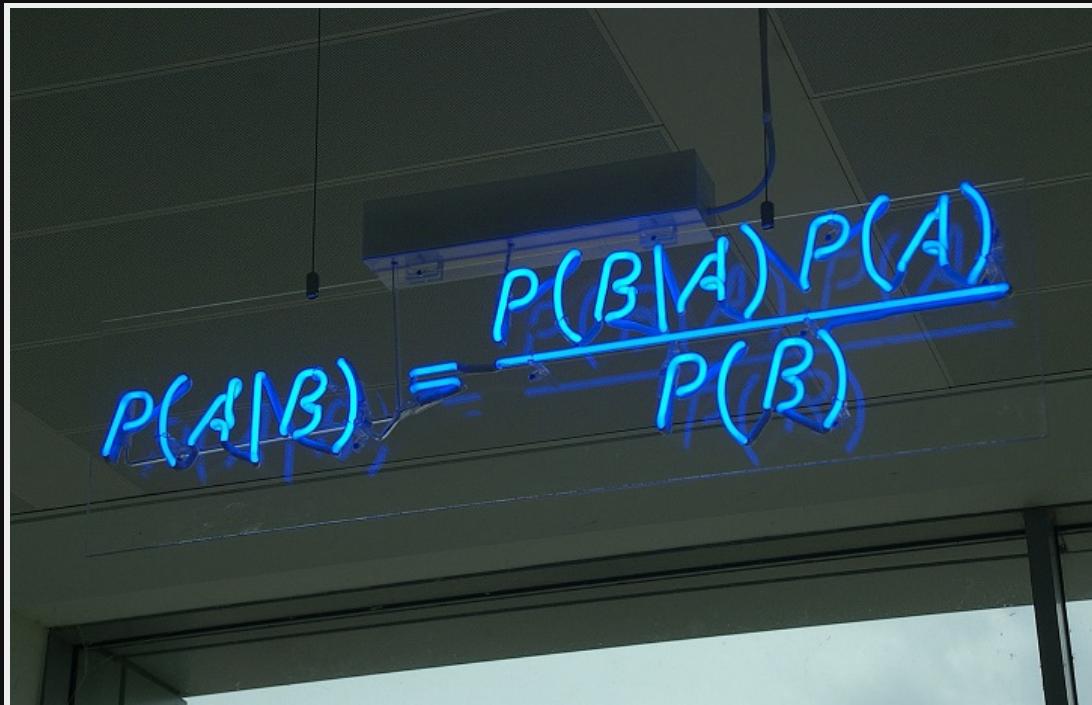
Precision & Recall

high precision, low recall

- **very careful**: doesn't return many positive answers, but if it does, it's mostly correct
- *Example*: rarely predicts spam; lots of spam emails end up in main folder

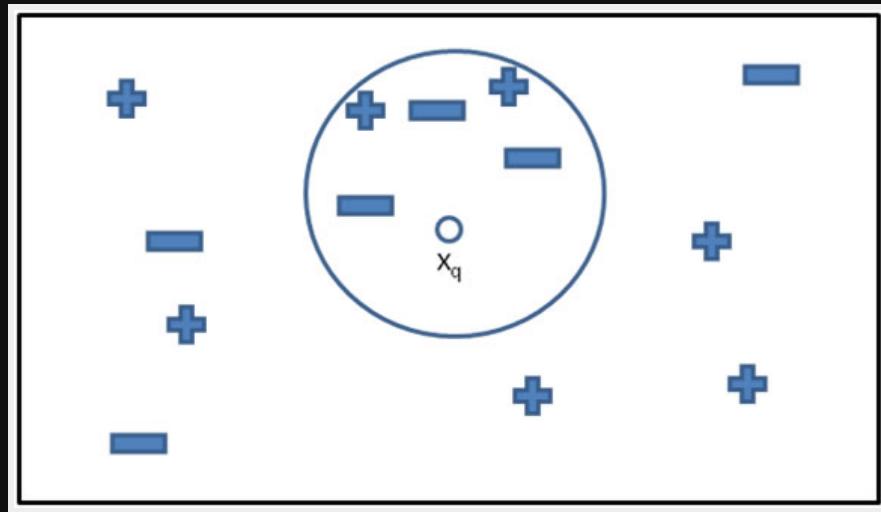
Machine Learning Algorithms

Naive Bayes



[<http://blog.yhat.com/static/img/bayes-theorem-neon-sign.jpg>]

K-Nearest Neighbor

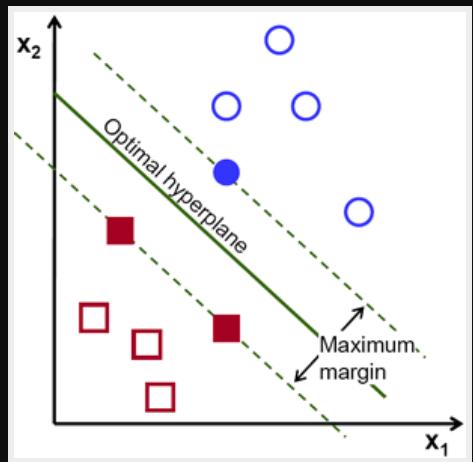


[<http://www.d.umn.edu/~deoka001/downloads/KNNExample.jpg>]

K-Nearest Neighbor

- L1 (Manhattan) or L2 (Euclidean) **distance**
- Does not train
- Not efficient in memory & computation

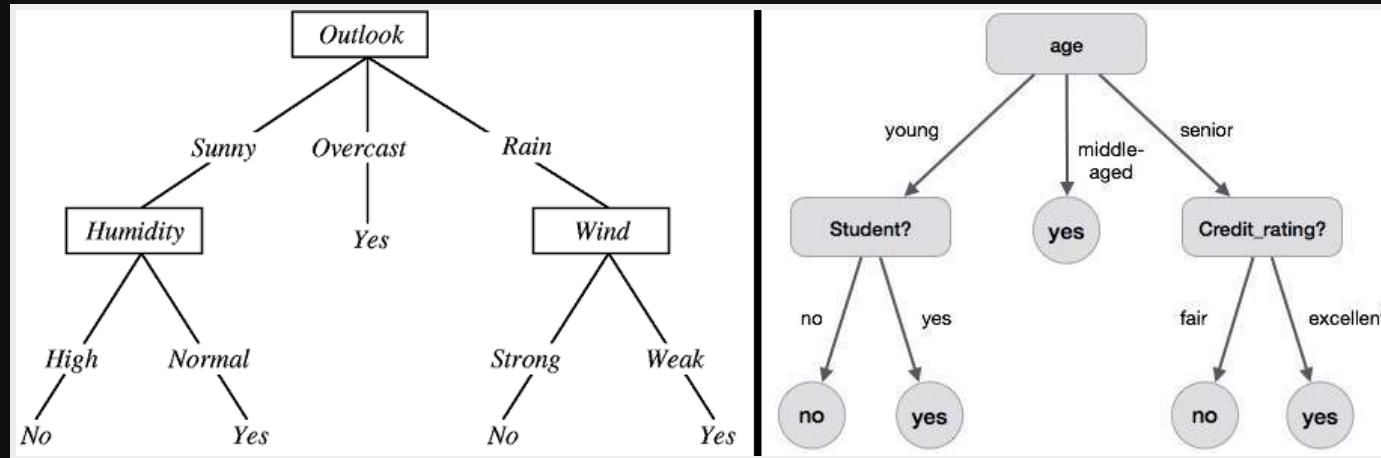
Support Vector Machines



[http://docs.opencv.org/2.4/_images/optimal-hyperplane.png]

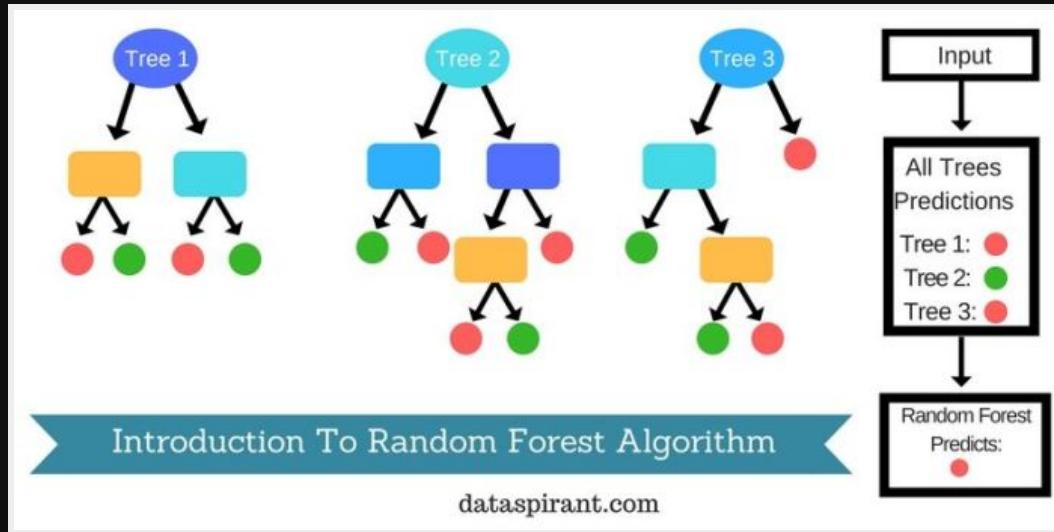
Decision Tree

Example: weather predictions



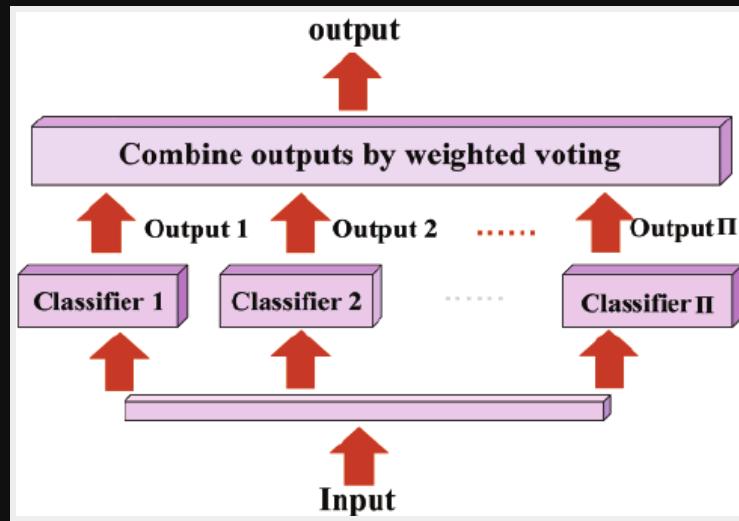
[<http://jmvidal.cse.sc.edu/talks/decisiontrees/dt.png>, https://www.tutorialspoint.com/data_mining/images/dm_decision_tree.jpg]

Random Forest



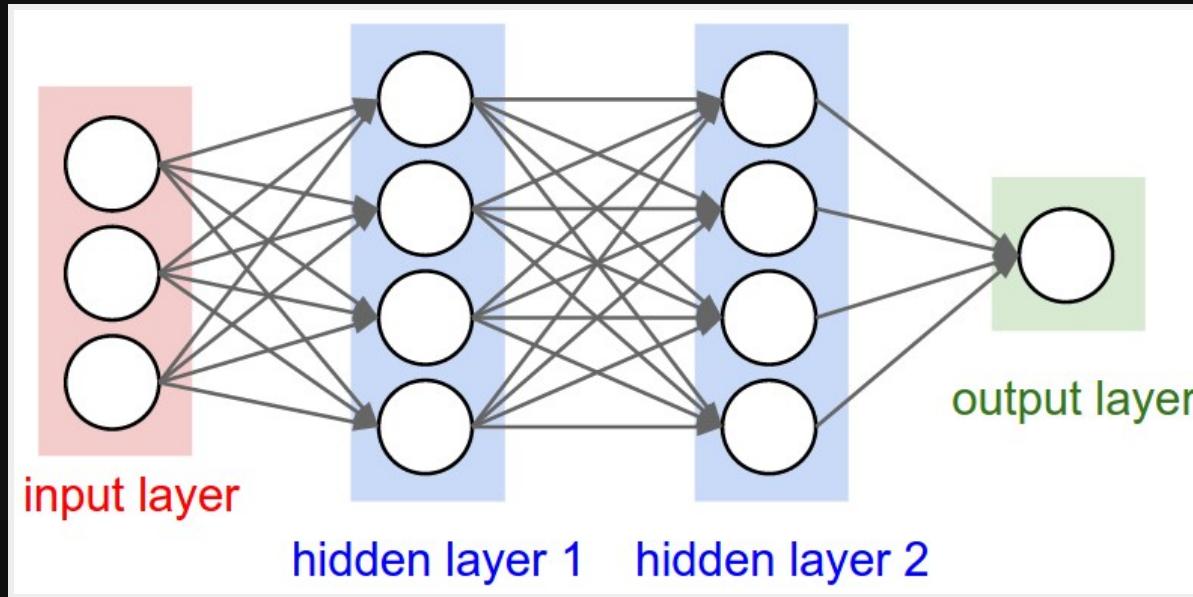
[<https://i1.wp.com/dataaspirant.com/wp-content/uploads/2017/04/Random-Forest-Introduction.jpg?resize=690%2C345>]

Ensemble



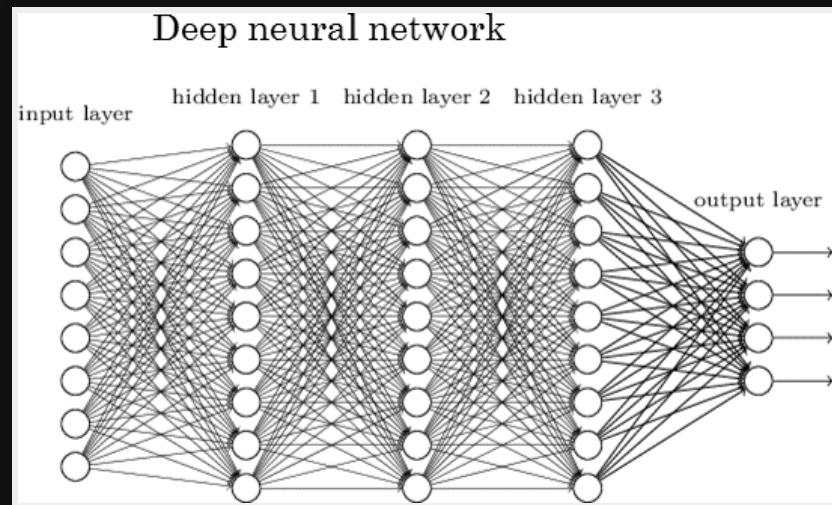
[https://www.researchgate.net/figure/6662177_fig1_Figure-2-Flowchart-to-show-how-the-ensemble-classifiers-NN-GO]

Neural Networks



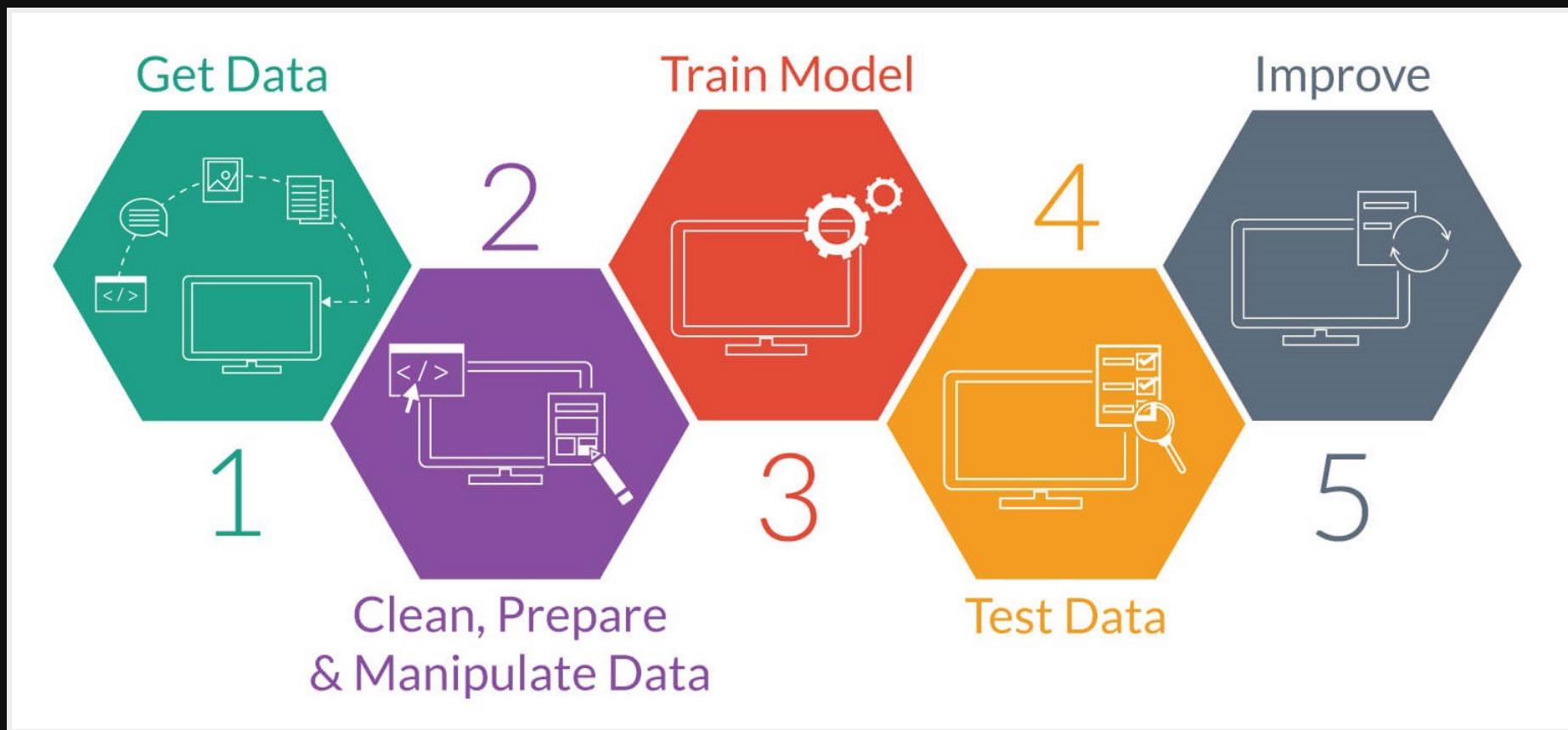
[http://cs231n.github.io/assets/nn1/neural_net2.jpeg]

Deep Learning



[<https://qph.ec.quoracdn.net/main-qimg-7c35987ad55173b3b76214b9112830ff>]

Summary



[https://cdn-images-1.medium.com/max/2000/1*KzmlIUYPmxgEHhXX7SlbP4w.jpeg]

References

[<https://www.coursera.org/learn/machine-learning>]

[<https://www.toptal.com/machine-learning/machine-learning-theory-an-introductory-primer>]

[<http://scikit-learn.org/stable/tutorial/basic/tutorial.html>]

[http://scikit-learn.org/stable/auto_examples/model_selection/plot_precision_recall.html]

[<https://medium.com/@ageitgey/machine-learning-is-fun-80ea3ec3c471>]

Announcements

- *Next meeting:* Classifiers
- MP2 will be release today
- Final Exam, instead of Final Project

Questions?