Practical aspects of Machine Learning

Piotr Mazur

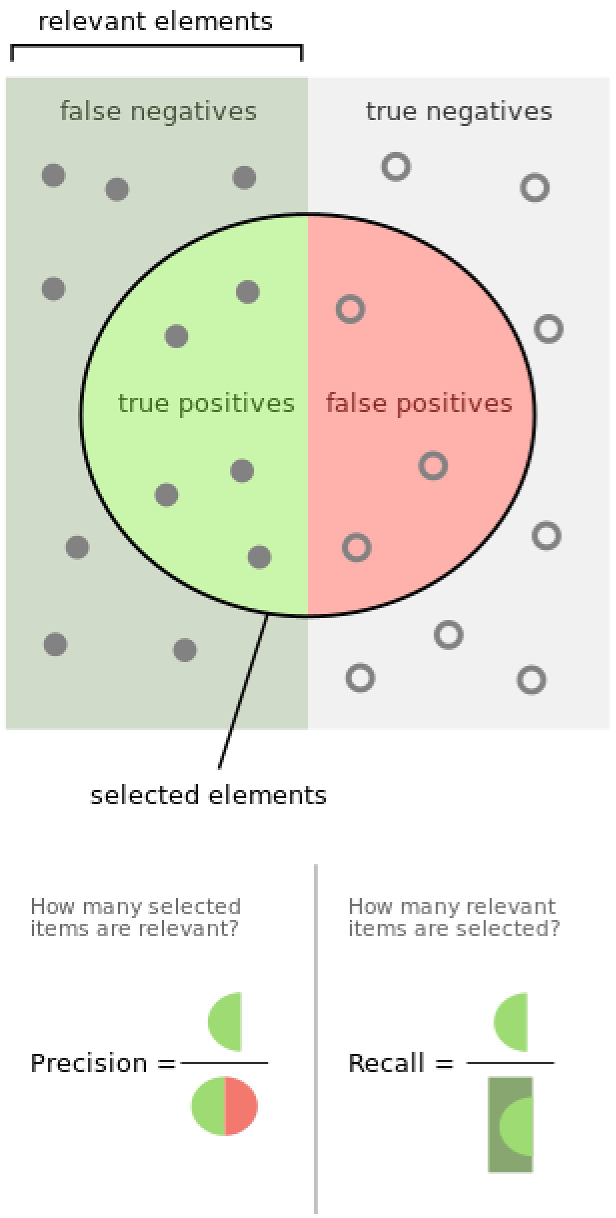
Agenda

- How to properly measure & improve our model?
  - Performance metrics
  - Regularization
  - Hyper parameter tunning
  - Transfer Learning
- ML project workflow
- How to keep up with state-of-art?
- Coding session

Performance Metrics

Classification Metrics

- Accuracy
- Precision
- Recall
- F1 score
- ROC
- AUC
- Confusion matrix
- And more…



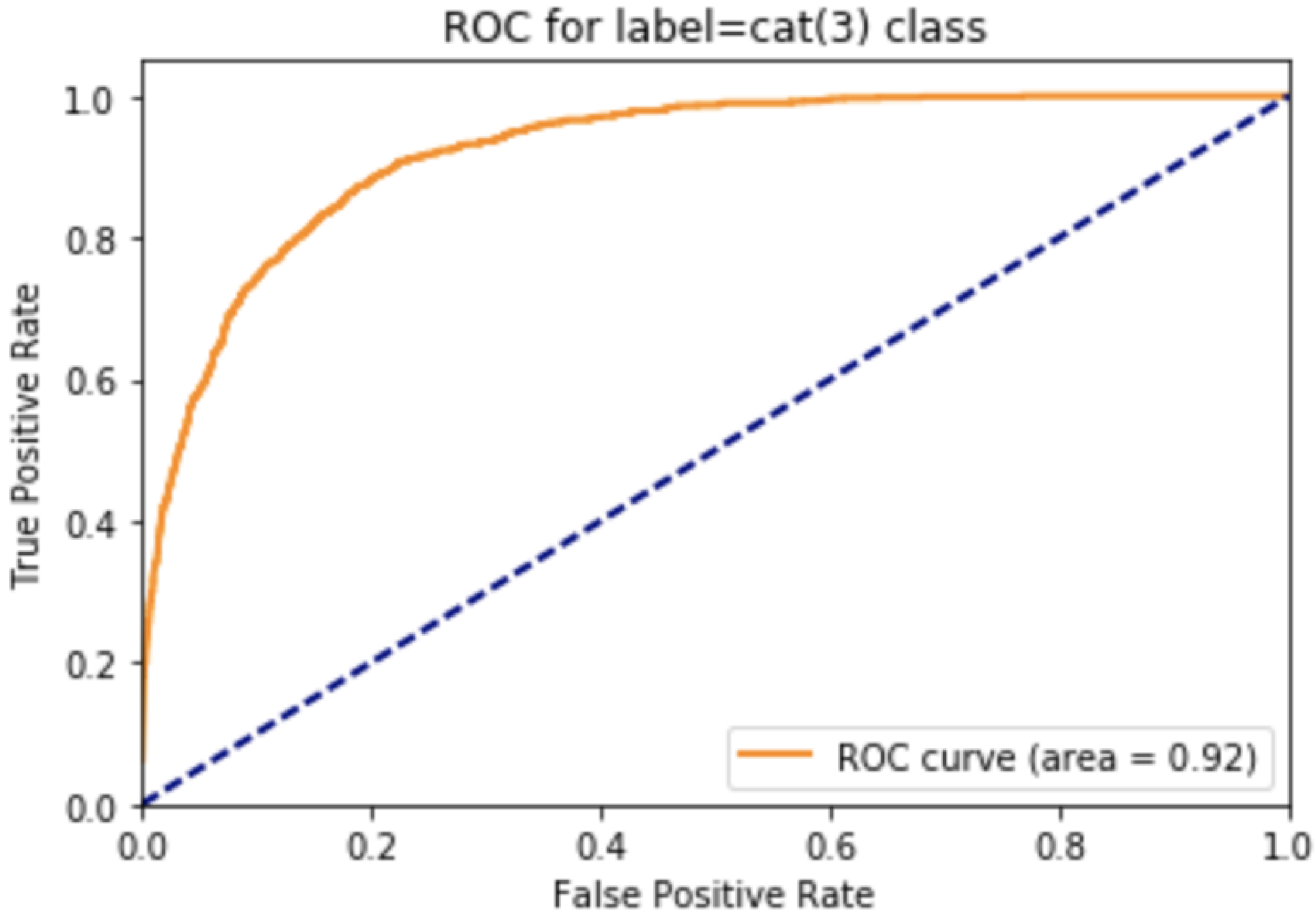
Source: https://en.wikipedia.org/wiki/Precision_and_recall

F1 score

$$F_1 = \left(\frac{2}{\mathrm{recall}^{-1} + \mathrm{precision}^{-1}} \right) = 2 \cdot \frac{\mathrm{precision} \cdot \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}}$$

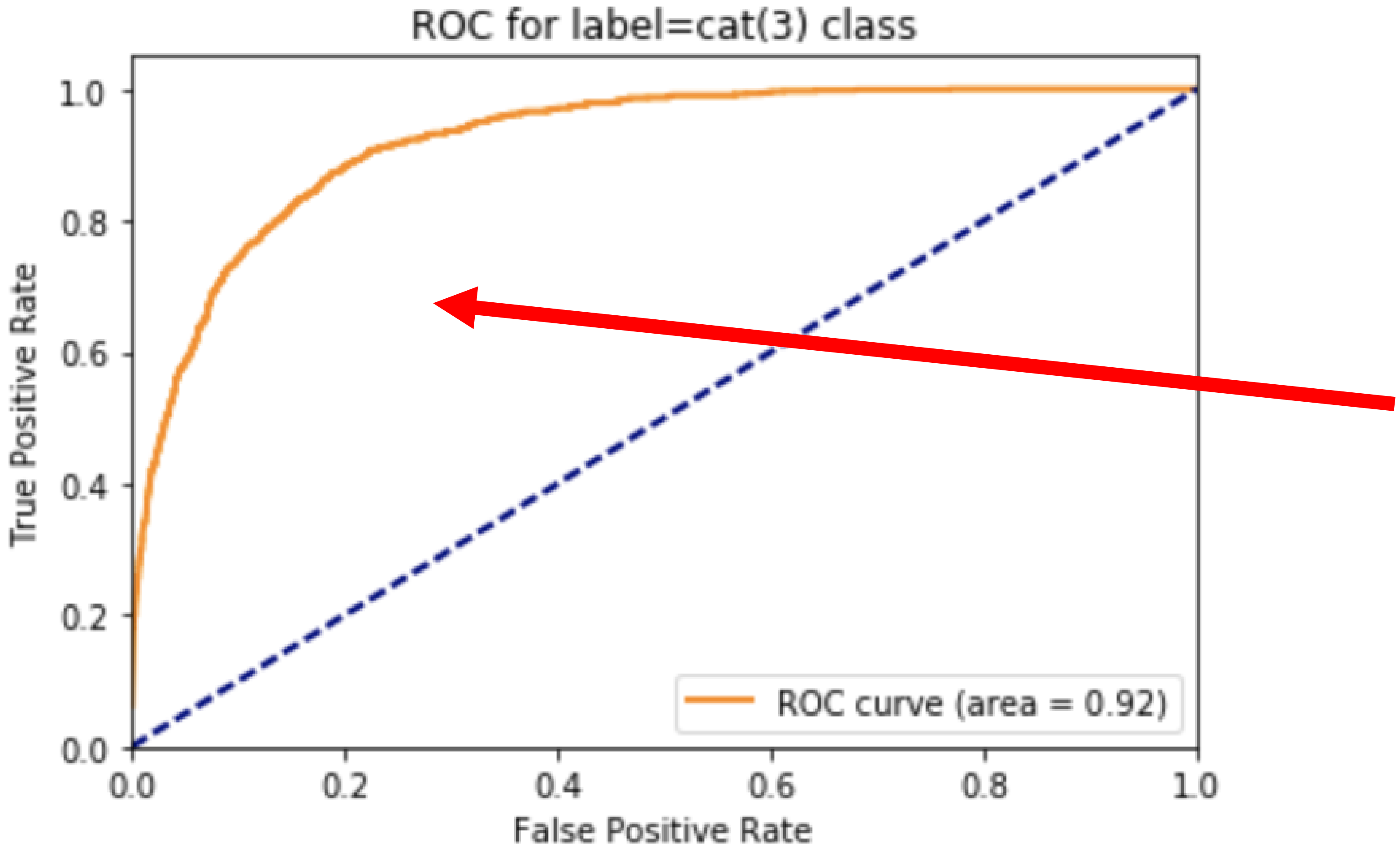
Source: https://en.wikipedia.org/wiki/F1_score

ROC



Source: http://bytepawn.com/solving-cifar-10-with-pytorch-and-skl.html

AUC



Source: http://bytepawn.com/solving-cifar-10-with-pytorch-and-skl.html

Confusion matrix

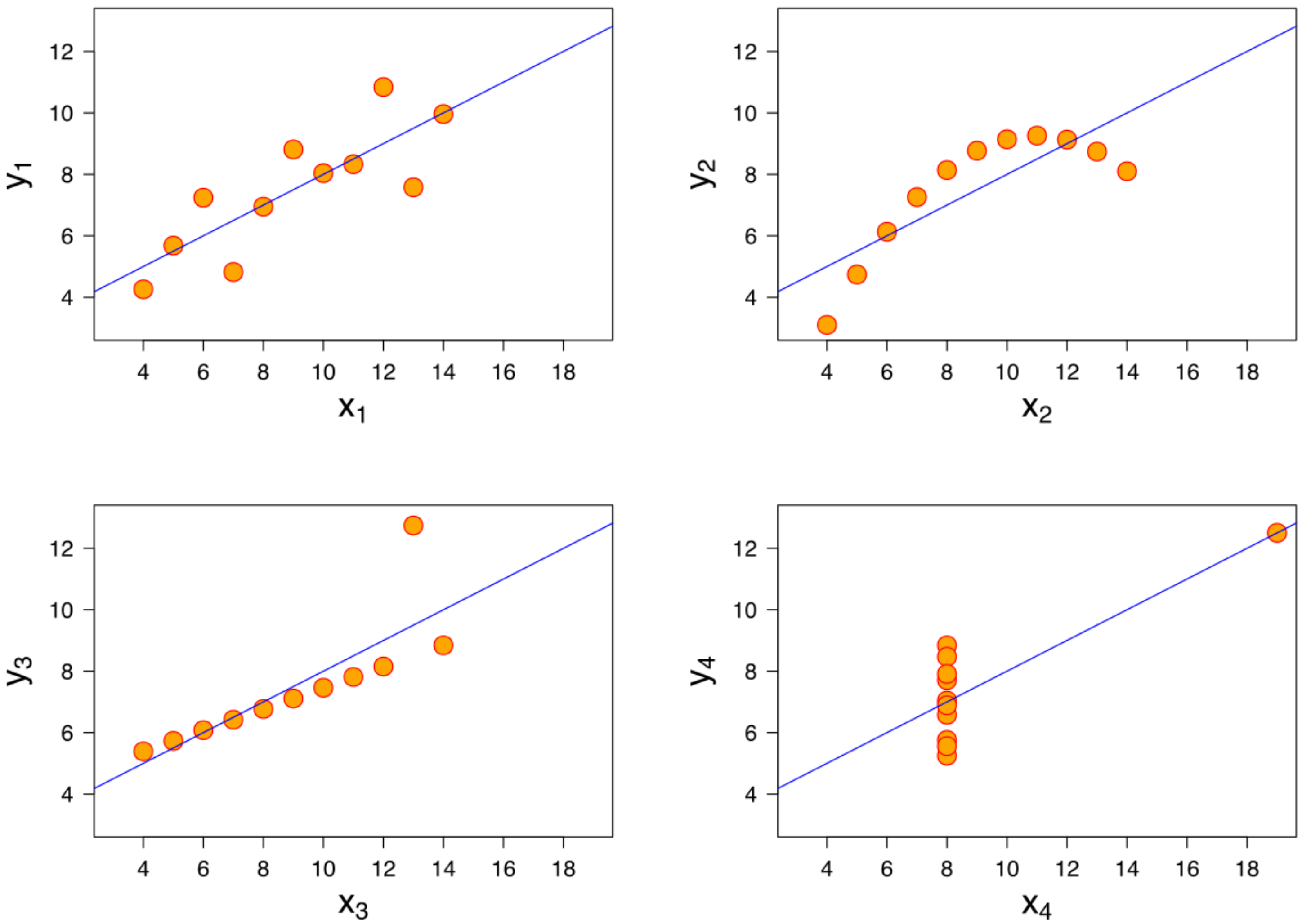
	predicted airplane	predicted automobile	predicted bird	predicted cat	predicted deer	predicted dog	predicted frog	predicted horse	predicted ship	predicted truck	recall
airplane	411	61	73	43	10	44	41	74	157	100	41%
automobile	37	434	41	37	23	18	93	43	65	222	43%
bird	66	43	287	61	67	65	222	73	44	24	30%
cat	30	64	120	214	28	190	212	64	39	56	21%
deer	38	31	205	50	177	59	273	94	28	42	18%
dog	20	38	131	153	25	295	225	61	43	34	29%
frog	5	21	121	99	34	47	571	33	11	38	58%
horse	26	33	128	75	79	59	123	348	24	82	36%
ship	142	88	39	18	8	37	21	21	489	140	49%
truck	45	187	38	41	7	17	69	64	80	474	46%
precision	50%	43%	24%	27%	39%	35%	31%	40%	50%	39%	accuracy 37%

Source: https://ml4a.github.io/demos/confusion_cifar/

Regression Metrics

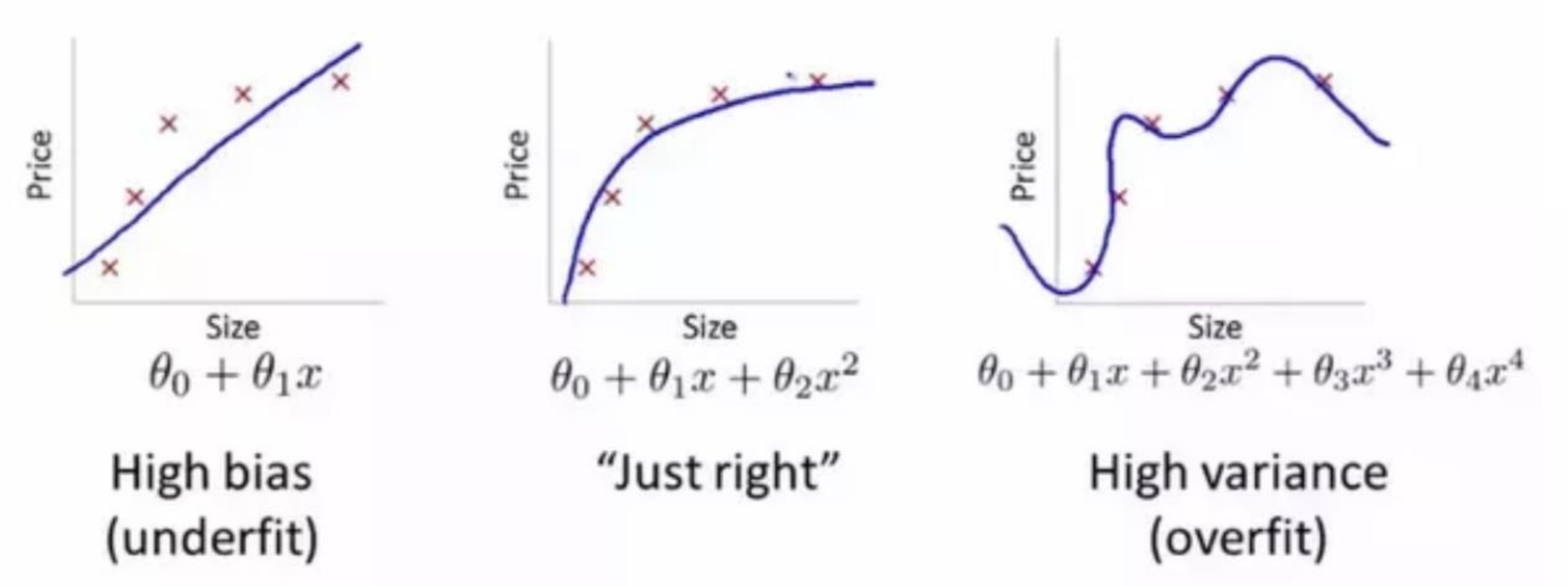
- Max error
- MAE - Mean absolute error
- MSE - Mean squared error
- R2 score
- And more…

R2 score



Source: https://en.wikipedia.org/wiki/Anscombe%27s_quartet

Bias-Variance Tradeoff



Source: Andrew Ng, Machine Learning Course on Coursera

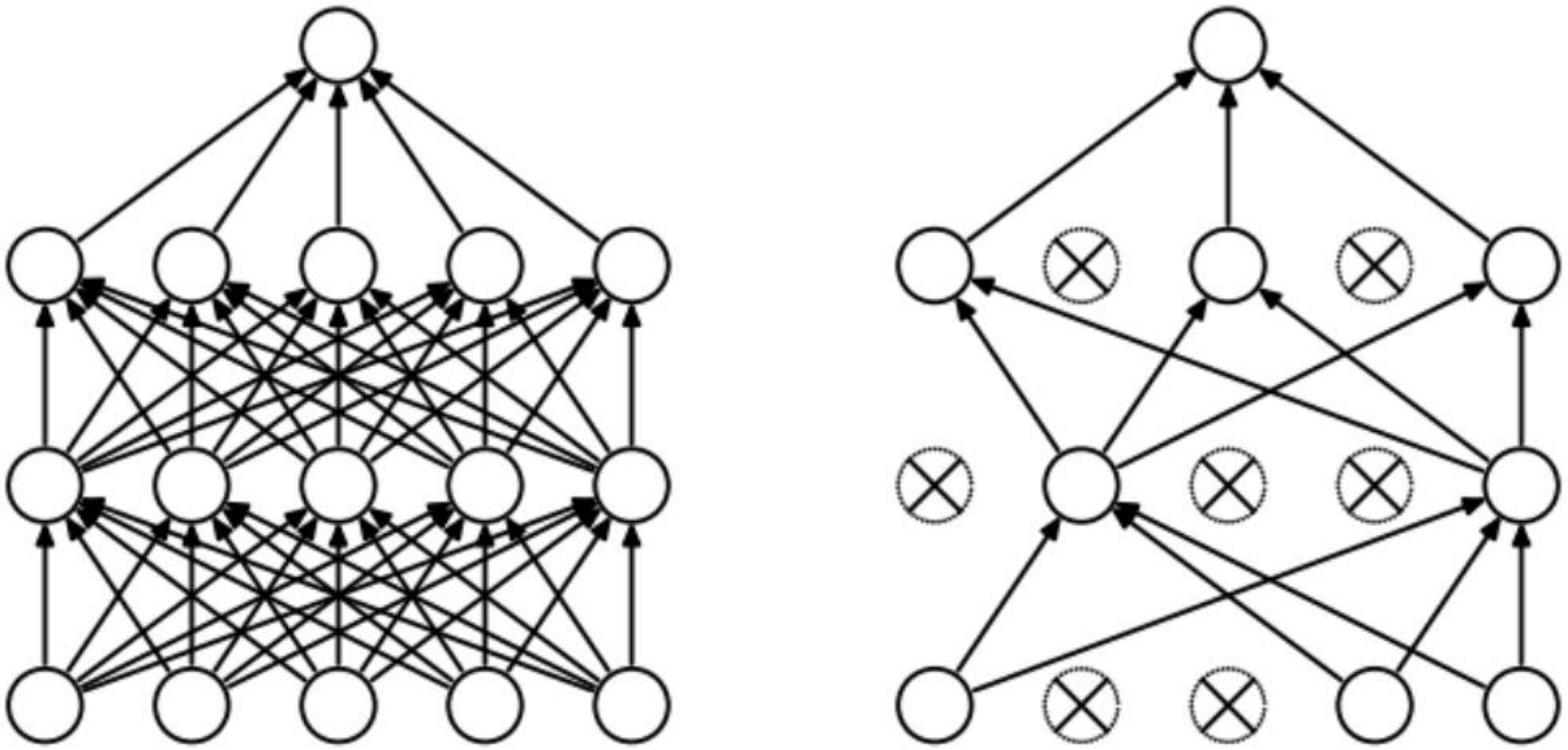
Regularization

What is the purpose?

Methods of regularization

- Dropout
- Early stopping
- Ensemble methods
- Weight decays
- Label smoothing
- Multi-task learning
- Data augmentation
- Adversarial training

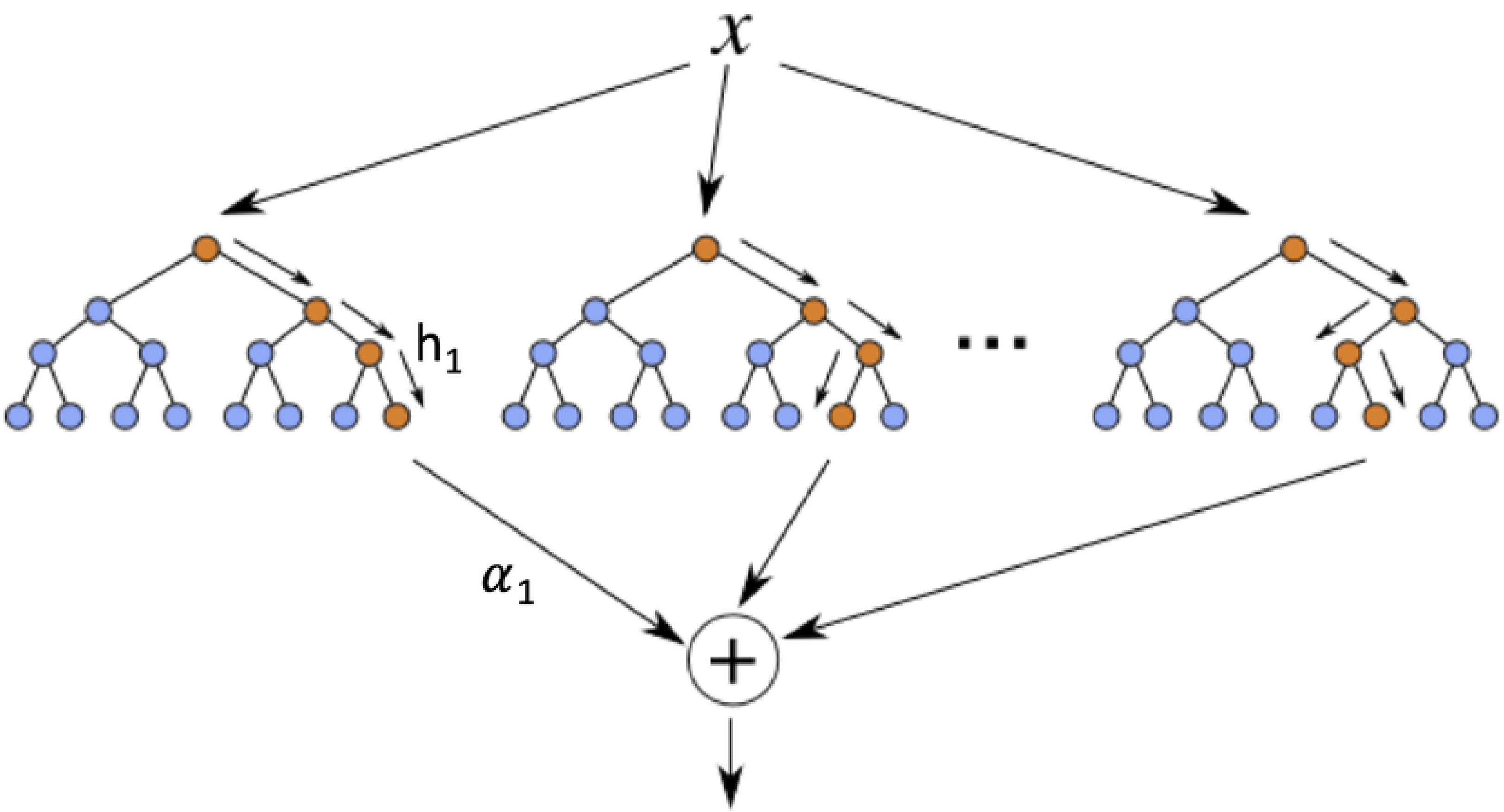
Dropout



(a) Standard Neural Net (b) After applying dropout.

Source: Srivastava, Nitish, et al. "Dropout: a simple way to prevent neural networks fromoverfitting", JMLR 2014

Ensemble Methods



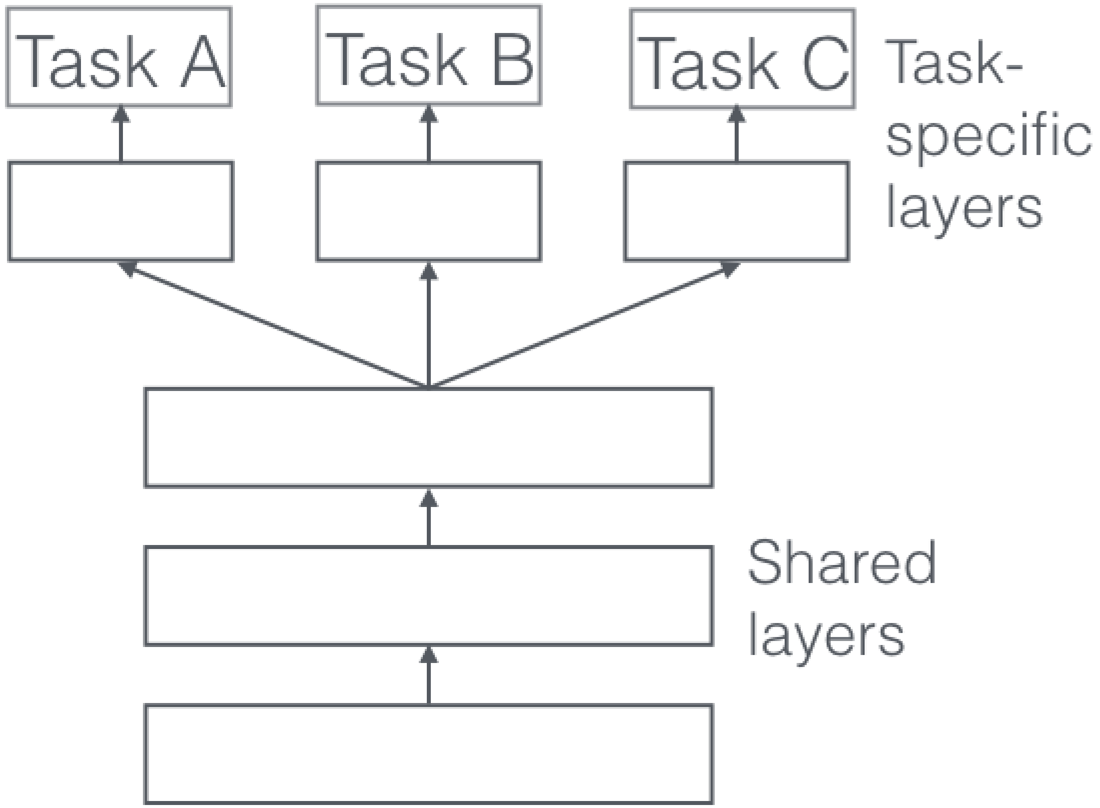
Source: https://www.researchgate.net/figure/Schematic-diagram-of-a-boosted-ensemble-of-decision-trees_fig2_325632132

Weight decays

Parameter norm penalties

Label smoothing

Multi-task learning



$$L_{total} = \sum_{t}^{|T|} \lambda_t L_t$$

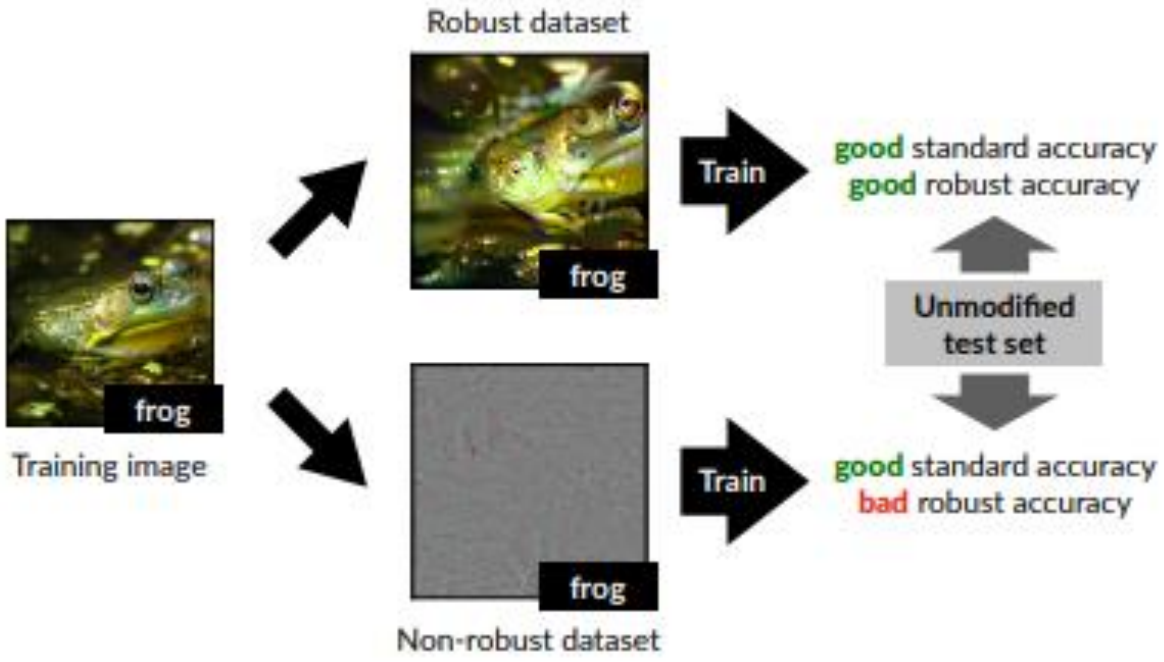
Source: https://towardsdatascience.com/self-paced-multitask-learning-76c26e9532d0

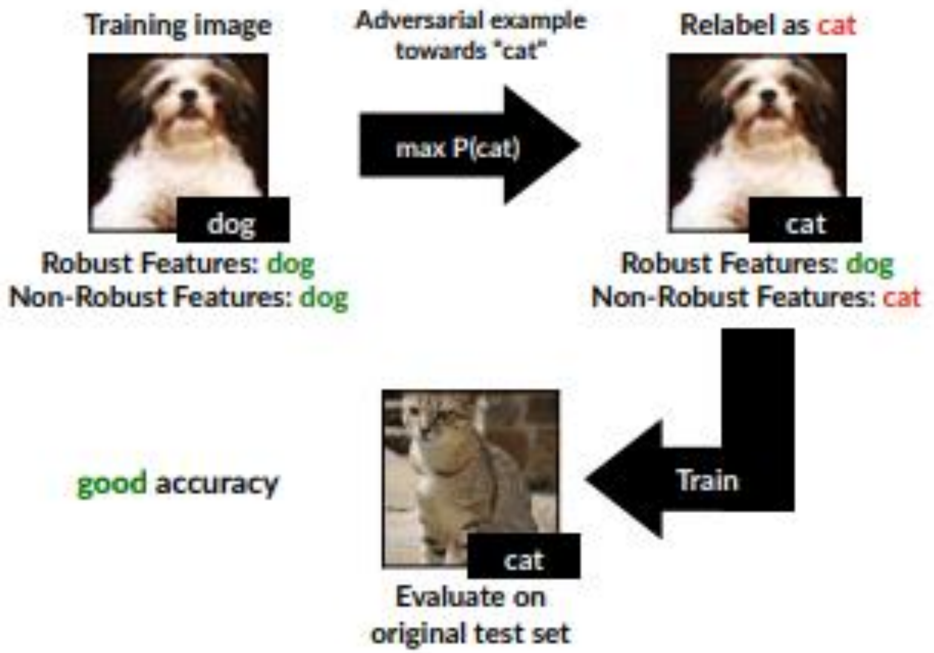
Data augmentation

Adversarial training

Source: https://medium.com/ai%C2%B3-theory-practice-business/adversarial-examples-are-not-bugs-585abca94d3c

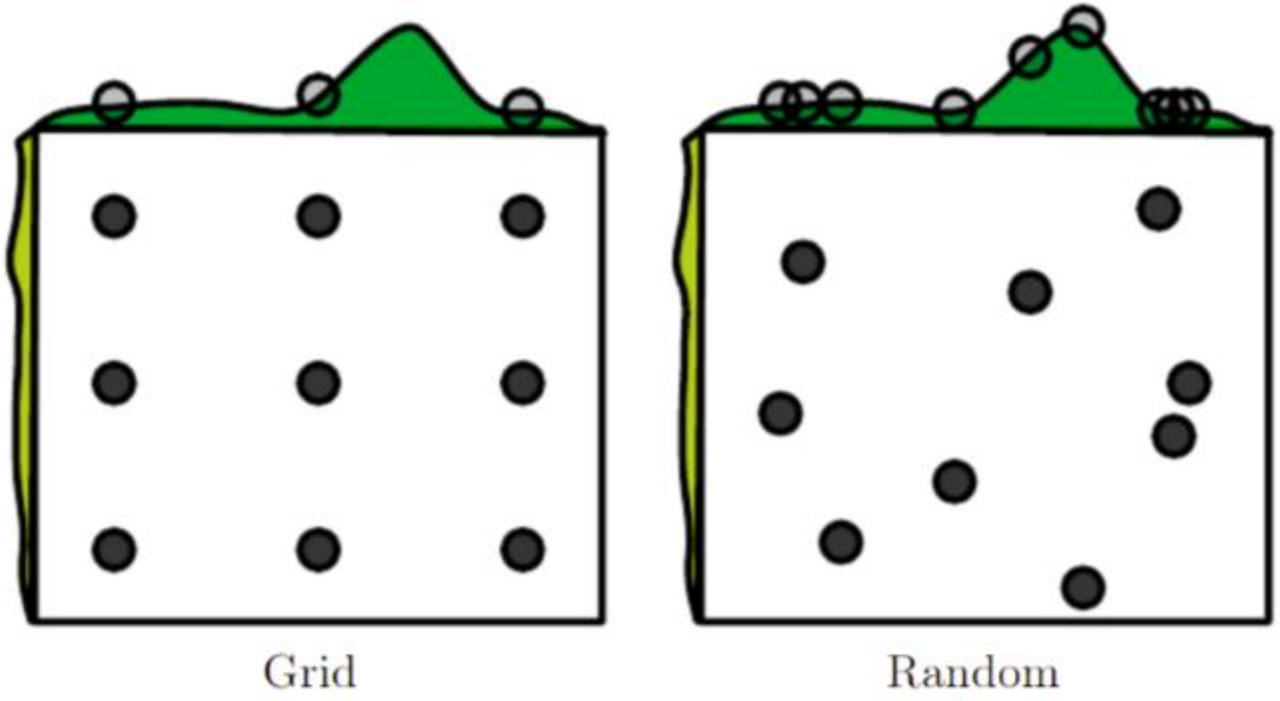
Adversarial training





Source: https://medium.com/ai%C2%B3-theory-practice-business/adversarial-examples-are-not-bugs-585abca94d3c

Hyper parameter tuning



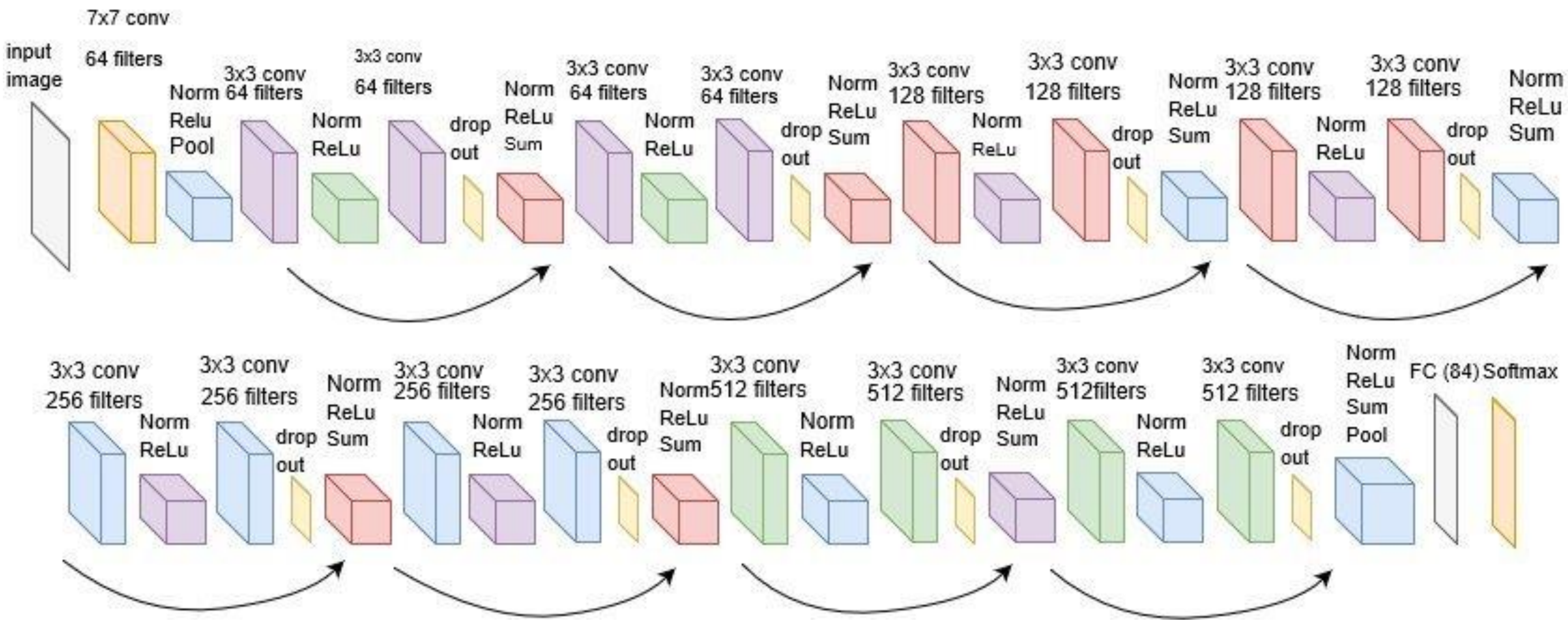
Source: Bengio Y. et al., „Deep Learning"

Transfer learning

Finetuning & fixed feature extractor

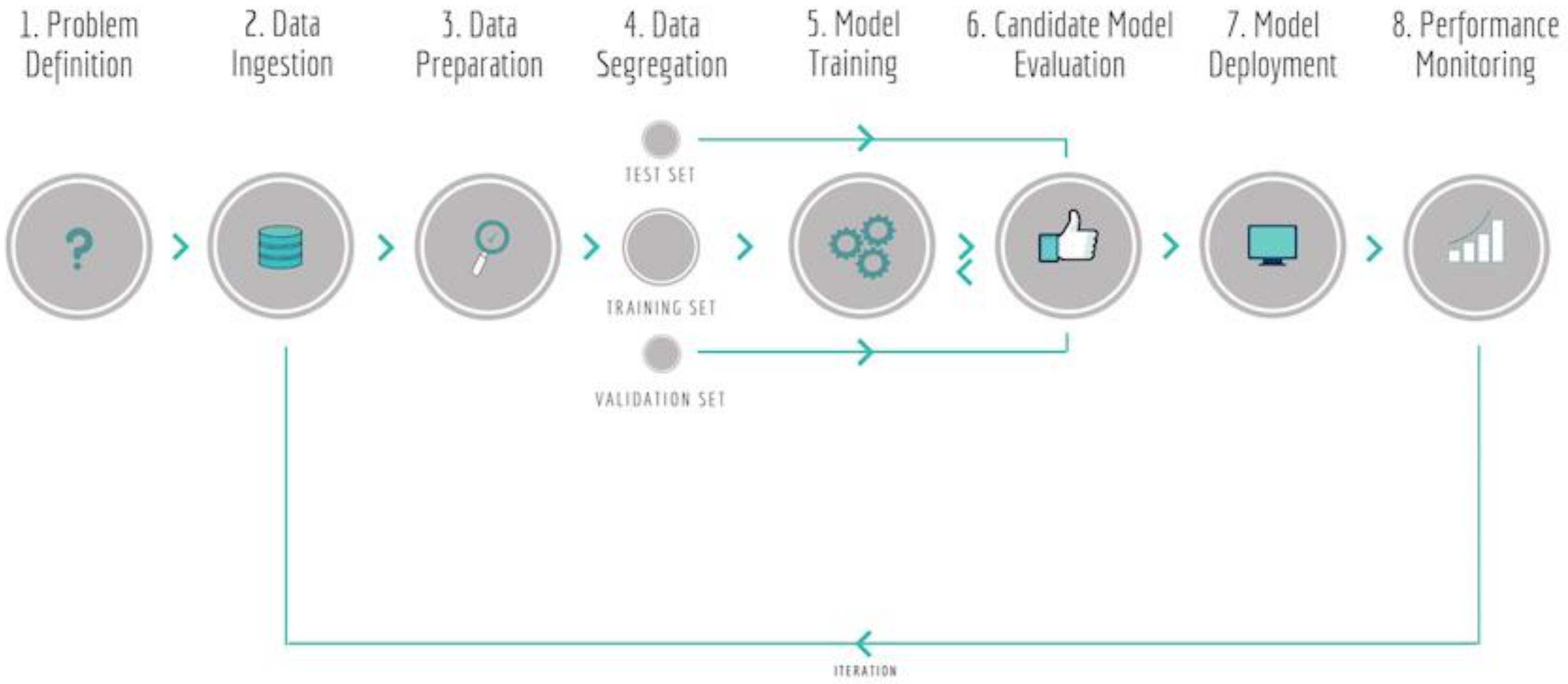
Coding

ResNet-18



7x7 conv
64 filters
input image
Norm Relu Pool
3x3 conv 64 filters
Norm ReLu
3x3 conv 64 filters
drop out
Norm ReLu Sum
3x3 conv 64 filters
Norm ReLu
3x3 conv 64 filters
drop out
Norm ReLu Sum
3x3 conv 128 filters
Norm ReLu
3x3 conv 128 filters
drop out
Norm ReLu Sum
3x3 conv 128 filters
Norm ReLu
3x3 conv 128 filters
drop out
Norm ReLu Sum
3x3 conv 256 filters
Norm ReLu
3x3 conv 256 filters
drop out
Norm ReLu Sum
3x3 conv 256 filters
Norm ReLu
3x3 conv 256 filters
drop out
Norm ReLu Sum
3x3 conv 512 filters
Norm ReLu
3x3 conv 512 filters
drop out
Norm ReLu Sum
3x3 conv 512filters
Norm ReLu
3x3 conv 512 filters
drop out
Norm ReLu Sum Pool
FC (84)
Softmax

Classic ML Workflow



Source: https://towardsdatascience.com/not-yet-another-article-on-machine-learning-e67f8812ba86

How my team won a hackathon

The challenge

The challenge

Challenge #3: Aviation challenge

Context

NAS (National Airspace System) of US is very dynamic and includes a lot of factors that might affect the airline's operations at network level and at the flight level. Those factors include weather, wind, airport capabilities and congestion in different parts of airspace. It is very difficult to plan efficient routes and be sure they will actually be flown.

Your task will be to help airlines predict the time of a flight.

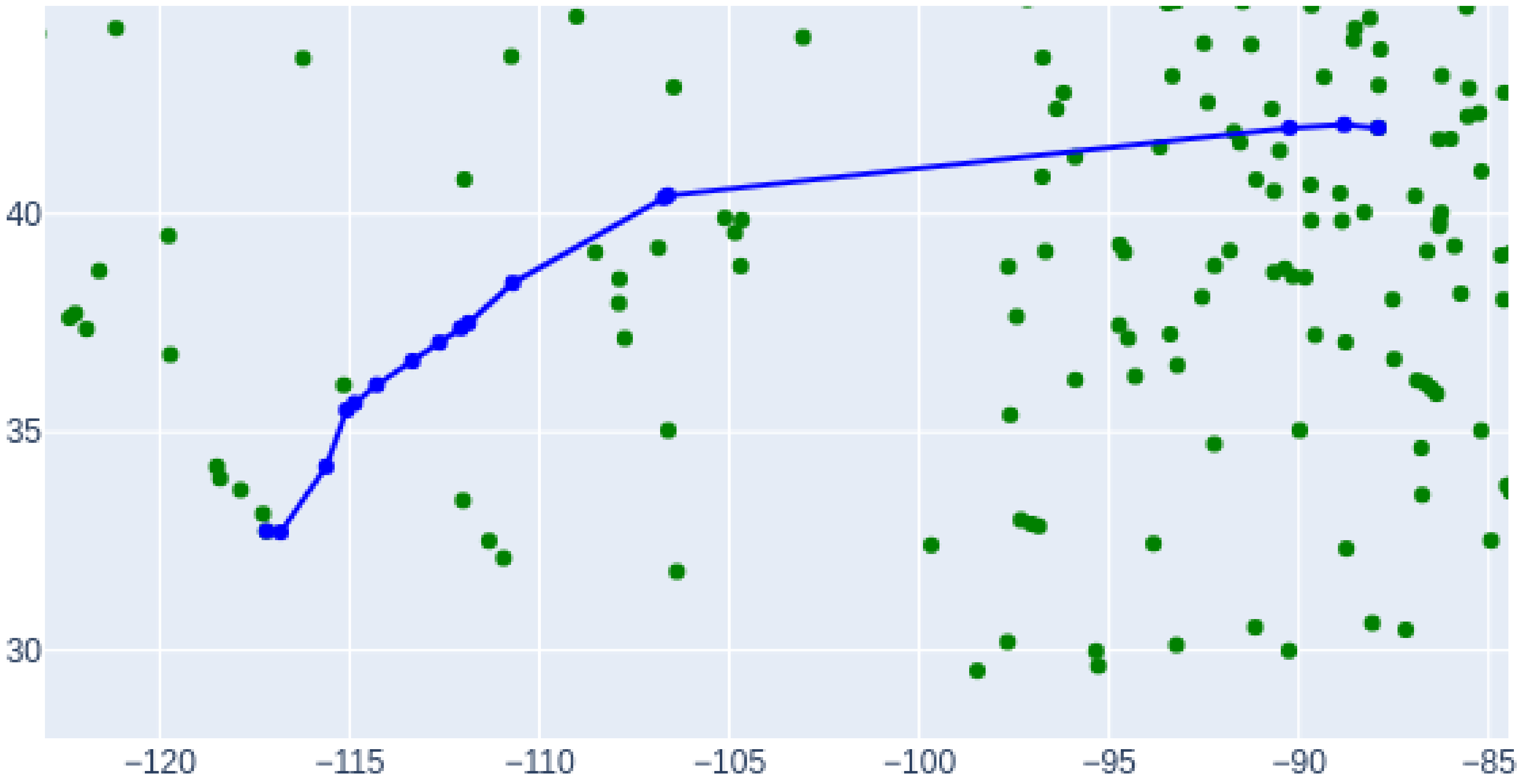
Baseline

MAE: 1023s

$$t = \frac{S}{V}$$

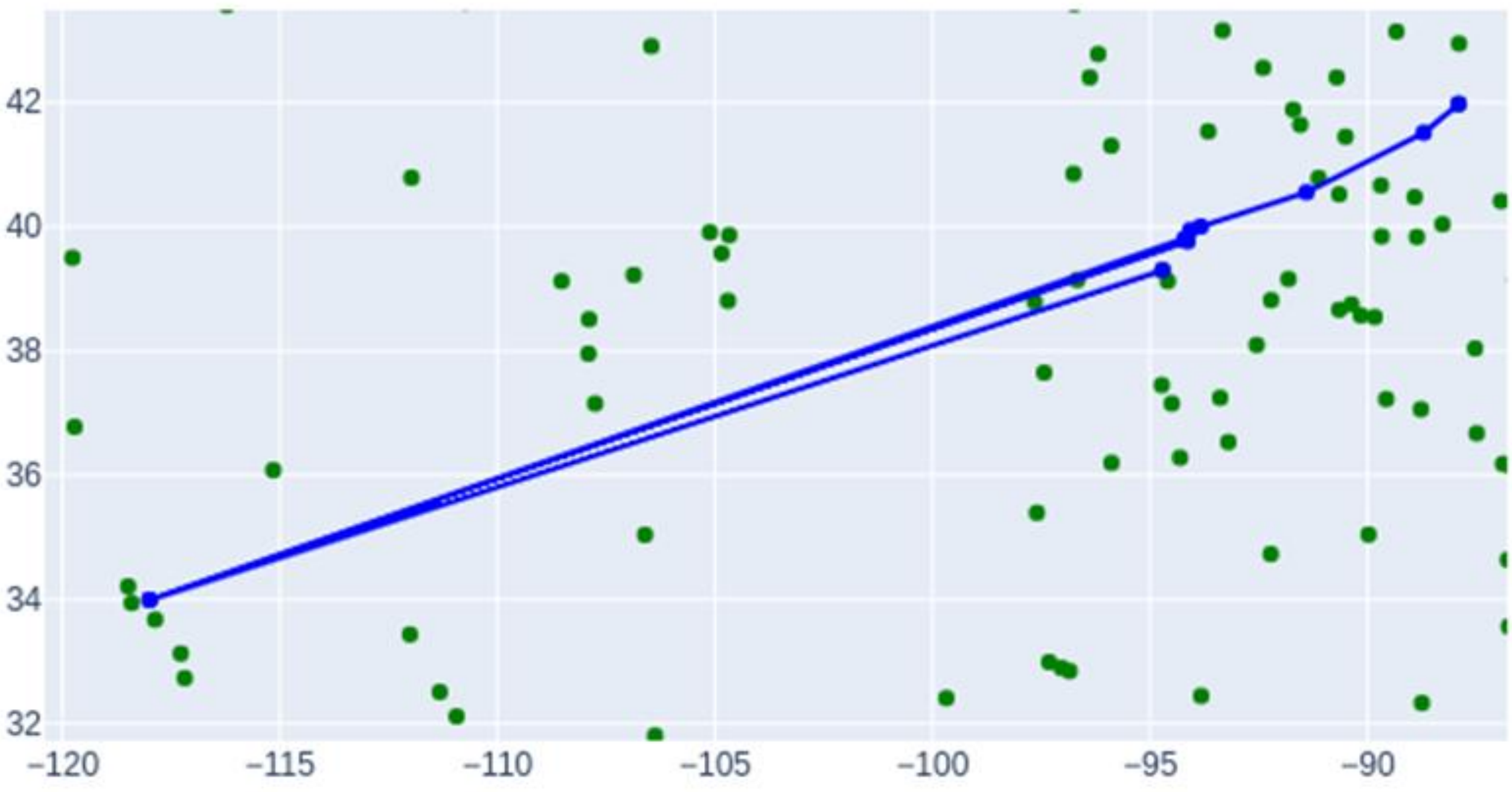
First data insights

MAE: 1023s



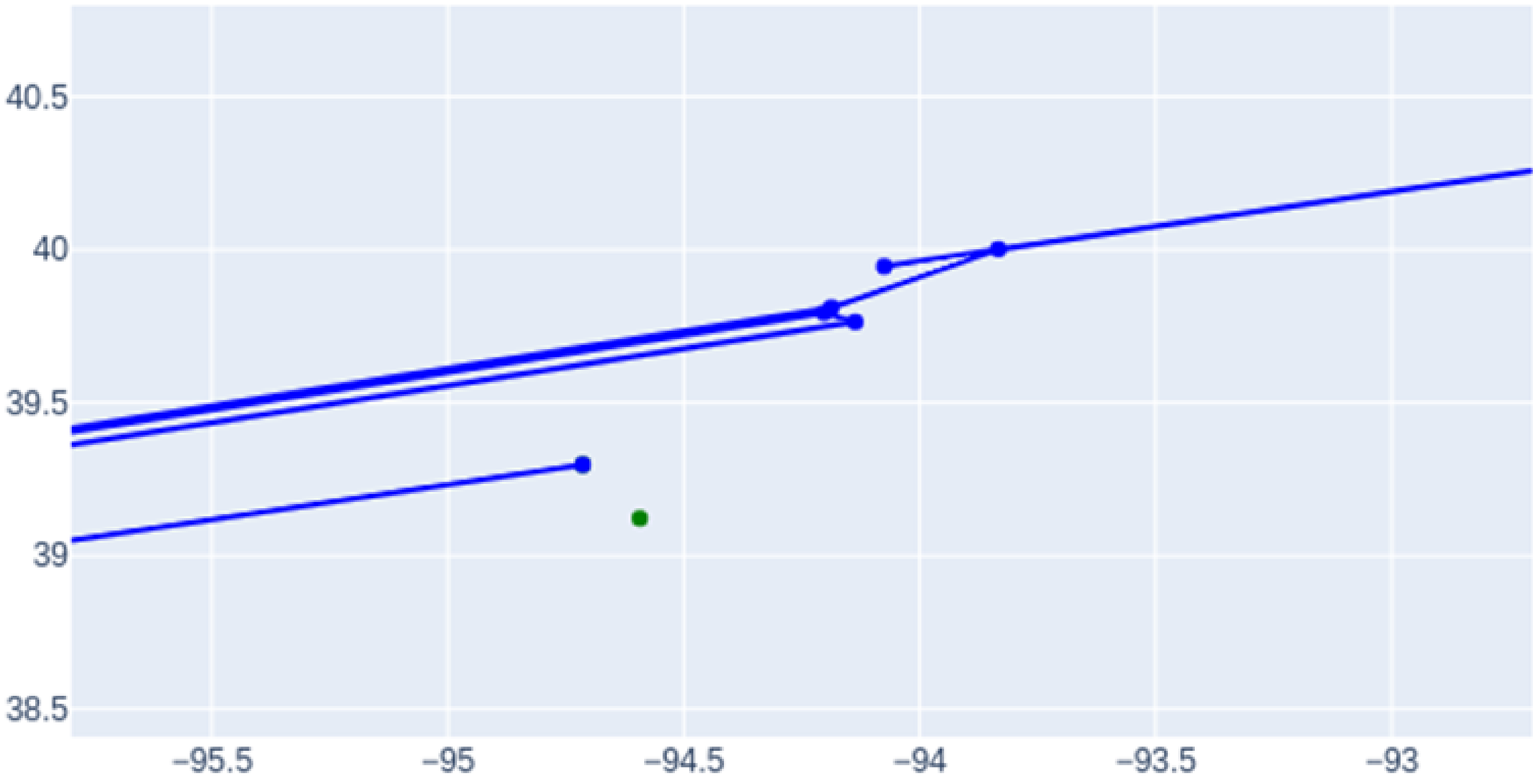
First data insights

MAE: 1023s



First data insights

MAE: 927s



Base model

MAE: 367s

- Linear Regression
- Gradient Boosting Regressor
- Random Forest Regressor
- Extra Trees Regressor

Features

MAE: 277s

- Cycles
  - Day of week
  - Day of month
  - Hour
- Airports
  - Coordinates
  - Angle
  - Distance
  - Top10 busiest airports indicator

MAE: 239s

Features

- Weather calculated at airports
  - Wind speed
  - Clouds
  - Vertically integrated liquid water
- Hyperparameter tuning
  - Extra Trees Regressor

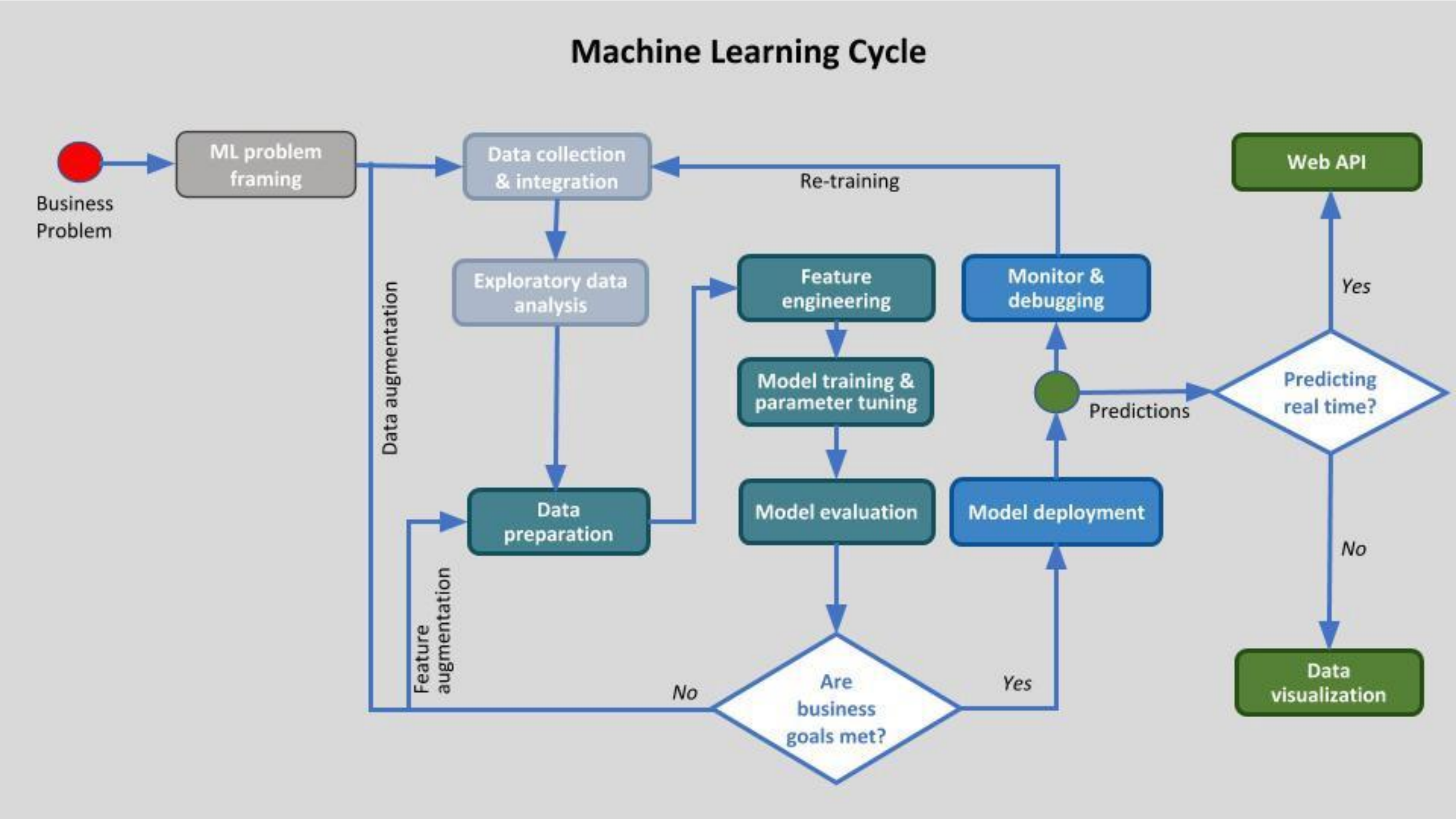
Results

1. 243s
2. 245s
3. 281s
4. 295s
5. 297s
6. 300s
7. 301s
8. 413s
9. 4814s
10. 5552s

Conclusions & possible enhancements

- More data
- LSTM
- Use weather from fixpoints
- Flight path analysis
- More information

ML Workflow



Source: https://medium.com/@kanjasaha/machine-learning-project-life-cycle-ea1e0bc8a2ed

How to keep up with state-of-art?

Influential people:

- Andrew Ng (Coursera)
- Andrej Karpathy (Tesla, formerly OpenAI)
- Demis Hassabis (DeepMind)
- Animashree Anandkumar (NVIDIA)
- Elon Musk (Tesla)
- Ian Goodfellow (Apple)
- Yann LeCun (Facebook)
- Geoffrey Hinton (Google Brain)
- Yoshua Bengio (University of Montreal)
- Fei-Fei Li (Stanford)

Conferences:

- ML in PL
- InfoShare
- NeurIPS
- PyData

PapersWithCode.com

https://paperswithcode.com/sota

Browse State-of-the-Art

1729 leaderboards • 1373 tasks • 1476 datasets • 20158 papers with code

Follow on Twitter for updates

Computer Vision

Semantic Segmentation

47 leaderboards

781 papers with code

Image Classification

71 leaderboards

659 papers with code

Object Detection

54 leaderboards

551 papers with code

Image Generation

63 leaderboards

267 papers with code

Pose Estimation

51 leaderboards

259 papers with code

See all 726 tasks

Natural Language Processing

Machine Translation

44 leaderboards

549 papers with code

Language Modelling

8 leaderboards

487 papers with code

Question Answering

46 leaderboards

464 papers with code

Sentiment Analysis

27 leaderboards

347 papers with code

Text Generation

28 leaderboards

203 papers with code

See all 271 tasks

Browse > Computer Vision > Image Classification

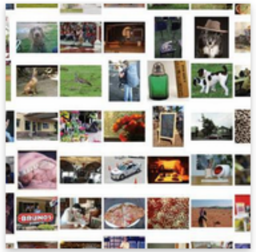


Image Classification

Edit

659 papers with code · Computer Vision

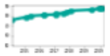
Image classification is the task of classifying an image into a class category. It is the most well-known computer vision task. Famous benchmarks include the MNIST dataset, for handwritten digit classification, and ImageNet, a large-scale image dataset for object classification.

Models are typically evaluated with an Accuracy metric, for example Top 1 and Top 5 Accuracy for ImageNet.

(Image credit: GAL: A Global-Attributes Assisted Labeling System for Outdoor Scenes)

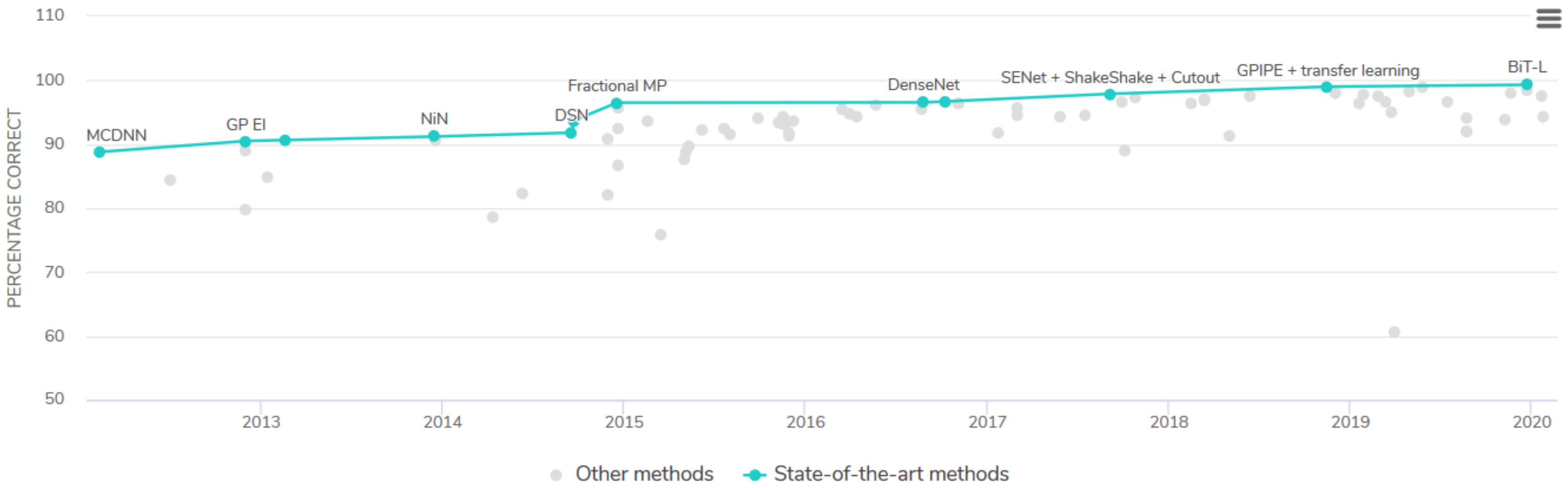
Leaderboards

Add a Result

TREND	DATASET	BEST METHOD	PAPER TITLE	PAPER	CODE	COMPARE
	ImageNet	NoisyStudent (EfficientNet-L2)	Self-training with Noisy Student improves ImageNet classification			See all
	CIFAR-10	BiT-L (ResNet)	Large Scale Learning of General Visual Representations for Transfer			See all
	CIFAR-100	BiT-L (ResNet)	Large Scale Learning of General Visual Representations for Transfer			See all

Browse > Computer Vision > Image Classification > CIFAR-10 dataset

Image Classification on CIFAR-10

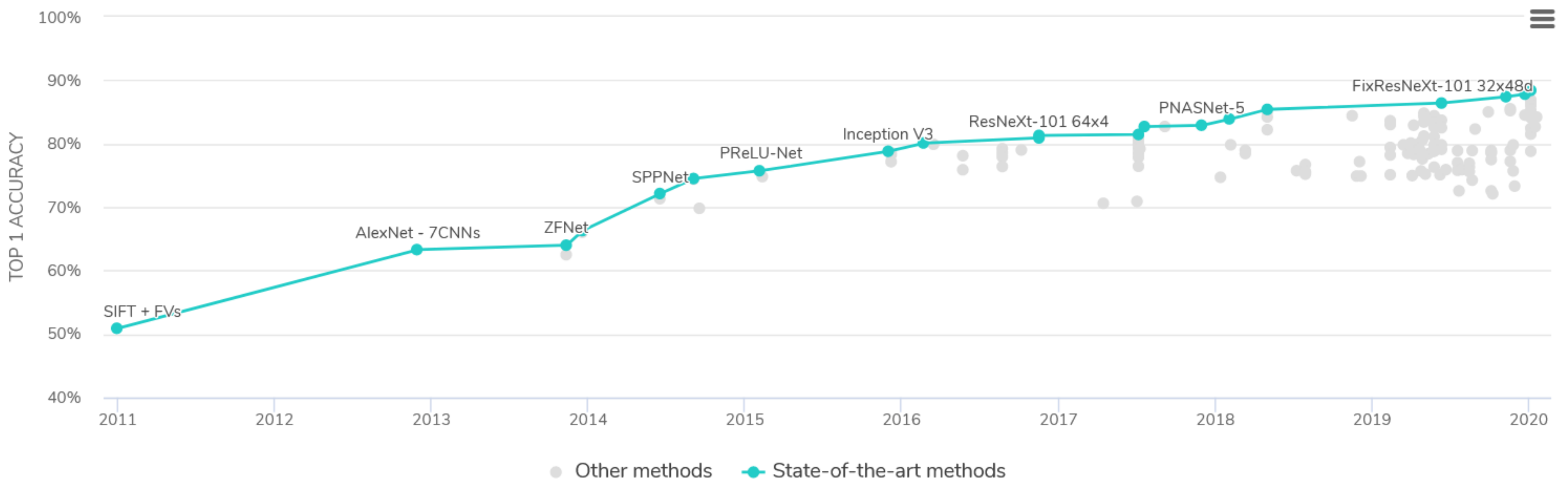


View: All methods

Edit

RANK	METHOD	PERCENTAGE CORRECT	PERCENTAGE ERROR	EXTRA TRAINING DATA	PAPER TITLE	YEAR	PAPER	CODE
1	BiT-L (ResNet)	99.3		✓	Large Scale Learning of General Visual Representations for Transfer	2019		
2	GPIPE + transfer learning	99	1	✓	GPipe: Efficient Training of Giant Neural Networks using Pipeline Parallelism	2018		

Image Classification on ImageNet



View: All methods

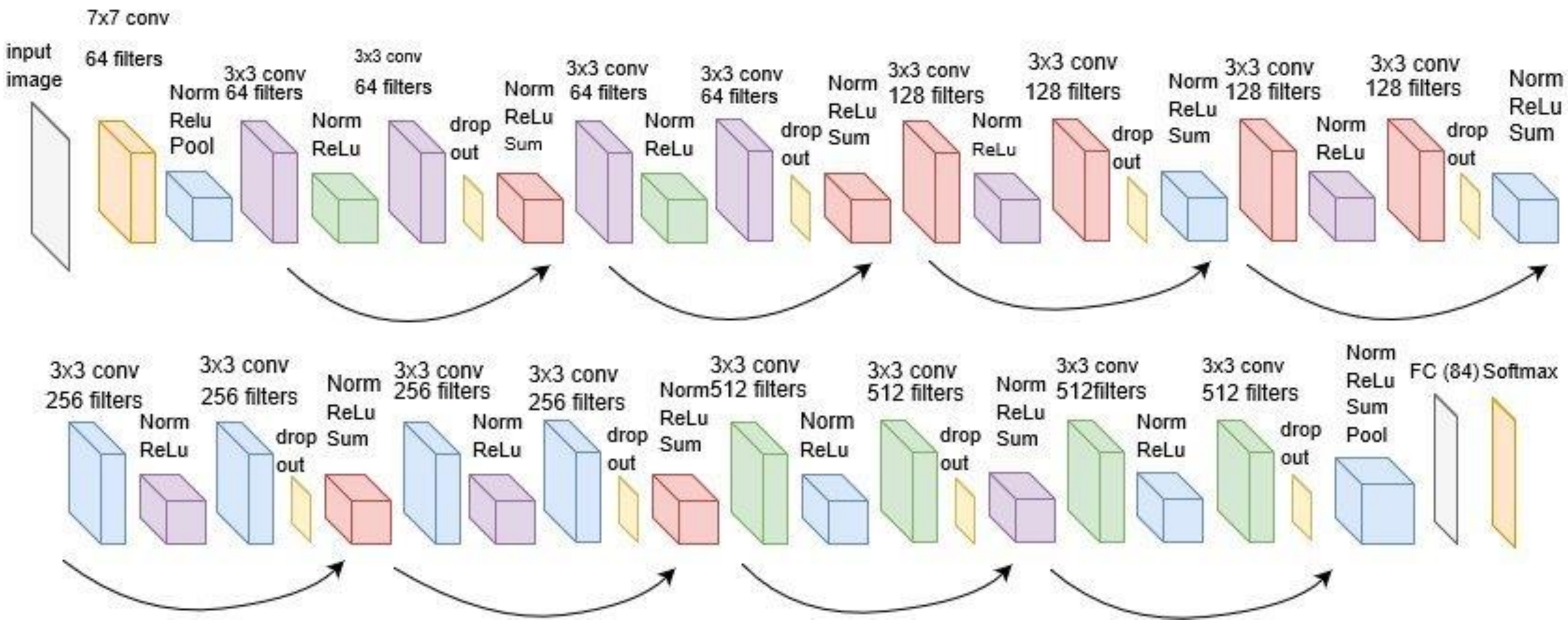
Edit

RANK	METHOD	TOP 1 ACCURACY	TOP 5 ACCURACY	NUMBER OF PARAMS	EXTRA TRAINING DATA	PAPER TITLE	YEAR	PAPER	CODE
1	**NoisyStudent** (EfficientNet-L2)	88.4%	98.7%	480M	✓	Self-training with Noisy Student improves ImageNet classification	2020		

Questions?

Coding

ResNet-18


input image
7x7 conv
64 filters
Norm
Relu
Pool
3x3 conv
64 filters
Norm
ReLu
3x3 conv
64 filters
drop
out
Norm
ReLu
Sum
3x3 conv
64 filters
Norm
ReLu
3x3 conv
64 filters
drop
out
Norm
ReLu
Sum
3x3 conv
128 filters
Norm
ReLu
3x3 conv
128 filters
drop
out
Norm
ReLu
Sum
3x3 conv
128 filters
Norm
ReLu
3x3 conv
128 filters
drop
out
Norm
ReLu
Sum
3x3 conv
256 filters
Norm
ReLu
3x3 conv
256 filters
drop
out
Norm
ReLu
Sum
3x3 conv
256 filters
Norm
ReLu
3x3 conv
256 filters
drop
out
Norm
ReLu
Sum
3x3 conv
512 filters
Norm
ReLu
3x3 conv
512 filters
drop
out
Norm
ReLu
Sum
3x3 conv
512filters
Norm
ReLu
3x3 conv
512 filters
drop
out
Norm
ReLu
Sum
Pool
FC (84)
Softmax

Questions?

Thanks for your attention

Piotr Mazur

piotrmmazur@outlook.com

kontakt@piomazur.pl

http://piomazur.pl/github

http://piomazur.pl/cnn-introduction

http://piomazur.pl/ml-practical